序

数学和任何一门自然科学或社会科学一样，其诞生和发展都离不开思维，都要依靠思维，并通过思维去研究和展现。对于作为思维存在形式之一的数学思维而言，人们也许会立即指出：其特征是高度的抽象性和严密的逻辑性，这固然言之有理，但却并不全然。随着数学方法论这一学科的诞生和发展，特别在我国数学教育界，自1980年以来在著名数学家徐利治先生的倡导下，数学方法论也日益受到重视和发展，以致人们在如下几个方面加深了认识。

其一是深刻认识到数学思维除了抽象性和逻辑性的特征外，同样具有甚或更重要的特征是观察、实验、类比、归纳、猜测、反驳、想象、直觉和美感等等；

其二是对于数学是一种文化这一古今有之的共识有了更深的理解，亦即进一步认识到数学不单纯是处理实际问题的工具，数学教育乃是一种文化素质的教育，或者说提高文化素质的原则，在数学教育中据有重要的位置。如所知，柏拉图曾张榜声明，不懂几何的人不要进他的哲学学校；英国律师至今都要在大学里学习许多数学知识，特别是许多高深的数学课程，都是被誉为西方名将摇篮之美国西点军校学生的必修课。其所以如此，无不源于提高文化素质的原则。几乎可以说，当这些学生真正成为哲学大师、著名律师或运筹帷幄的将帅时，实际上早把学生

时代所学到的那些具体的数学知识忘得一干二净，但他们在当年所受到的数学训练，却一直在他们的工作和事业中起着重要作用，甚至受用终身；

其三是进一步认识到任何重大数学发现和创新，都是数学思维方式发生变革的结果。如所知，大数学家希尔伯特曾指出：任何历史遗留难题的解决，或者任何重大的开创性数学工作的出现，都将伴随着新的数学园地的开辟。这无非是说，只有那些囿于传统思维方式和传统观念统治下无法解决的问题才能长期不得其解，因而此类问题的解决或其他有重大历史意义的开创性工作，都要有新的思维方式的出现和对于传统观念的突破，从而也就必然随之而开辟新的数学研究方向和领域。既然如此，我们若能深入探索数学思维方式的变化规律，进而自觉地把握和运用这些数学思维方式的变化规律，其意义之重大是不言而喻的。

总之，我国数学教育界已日益重视数学创造性活动中的数学形象思维与数学思维方式的研究，有关文献和著作日益增多，国内部分高等院校相继开设数学方法论课程，成立相应的研究中心；培养了一批数学方法论方向的研究生，多次举办全国性的数学方法论学术讨论会。席振伟所著《数学的思维方式》一书，无疑又是日益增多之数学方法论文献、著作中的一支奇葩。该书首先从哲学、现代心理学和思维科学的角度论述了思维概念的本质和特征，进而论述了数学的若干基本思维方式的含义、历史渊源、方法论功能及其应用，再则论述数学中的具体思维原理和方法。这也是本书作者多年来从事数学教学与科研的一项重要成果。全书内容丰富，说理清楚，堪称我国数学教育和数学方法论方面的又一部佳作，因而本人乐

于向读者推荐并为之作序。本书读者们有何积极建议或评述，可与本书作者通信讨论，以求更加完善和达到更高的学术水平。学无止境，相信作者会热烈欢迎和衷心感谢的。

朱梧槚

1993年12月9日于南京

出版说明

如大家所知，数学方法论作为研究数学中的发现、发明与创新等法则的一门学问，已有很长的历史，而且内容极为丰富。16世纪以来，如笛卡尔(Descartes)、莱布尼兹(Leibniz)、庞加莱(Poincaré)、克莱因(Klein)、希尔伯特(Hilbert)和阿达玛(Hadamard)等著名学者，都有过这方面的论著和发表过这方面的精辟见解。就近现代而言，以著名的美籍匈牙利数学家波利亚(Polya)为例，他曾以数十年的时间从事数学方法论的研究，出版了一系列论著，并被译为多种文字，受到全世界的普遍重视，被誉为第二次世界大战后出现的经典著作之一。在我国，也有许多学者在各种不同的场合屡次指出：要在数学教材与教学过程中，注意对形成数学概念的认识过程的分析，努力教给学生以寻找真理和发现真理的手段，特别是我国数学家徐利治教授，他先后到过苏联、德国、美国、加拿大和保加利亚等国进行学术交流，结合国内实际情况研究了世界数学的历史和现状，深感在教学与科研领域中，有大力提倡数学方法论的必要。在他的倡议下，我国一些理工科大学和师范院校相继开设了数学方法论选修课，出版界也出版了一些这方面的专著和通俗读物，这无疑是一个令人鼓舞而又富于开创性的发展趋势。然而总的说来，在现今的数学教育与数学教学过程中，主要的倾向还是偏重逻辑思维能力的训练，对于如何教给学生以寻找真理和发现真理的本领不够重视，在一定程度上低估了发散思维的训练在智力开发中的作用，以致不能较好地培养学

生的创造能力。

上述情况表明，我们仍需大力提倡数学方法论的研究，并应把数学方法论应用到中学与大学的数学教育实践中去。特别是，我国现今正处在四个现代化建设和数学教学改革的新时期，这就急需培养出一支高水平的、庞大的科技队伍，而尤其急需造就一支高水平的、庞大的数学教师队伍，因为这是我国能否建成科技大国的关键。正是为了适应这一形势的需要，我社自1986年初就开始酝酿和筹备出版《数学方法论丛书》（以下简称《丛书》），并拟请徐利治教授主持此项工作。此举得到了当时正在美国访问讲学的徐利治教授的赞同。全国各地的有关专家、教授也很支持此项工作，纷纷承担《丛书》编写任务。1987年4月，我社与徐利治教授等充分磋商，组建了《丛书》编辑委员会与特聘顾问。我们深信，在《丛书》的全体编委的共同努力下，一定能在高水平和高质量的基础上出版好这一套《丛书》，我们也由此而希望，这套《丛书》的出版，能在我国数学教学改革和培养人才的事业中有所贡献。

《丛书》共分三个档次，除了少数几本属于高档次的专著之外，其他两个档次主要面向中学教师、大专院校学生、研究生和一般数学爱好者。无疑，《丛书》中的大部分题材，对于使用数学工具的科技工作者来说也是有启发性的。

限于水平，在《丛书》的编辑和出版过程中，难免会有缺点和差错。热切希望数学教育界人士和广大读者多多批评指正。

江苏教育出版社

1988年6月

目录

一 思维与数学思维

对数学思维方式的研究与探索，是以思维和数学思维这两个基本概念作为基础和前提的。

本章将从思维的概念入手，论述数学思维的特性、品质、基本成分和思维结构，并进而揭示作为数学思维结构重要维度的数学思维方式的含义、内容及其对数学思维能力培养的重要意义。

1.1 思 维

1. 思维的概念与实质

思维从不同的角度观察有着不同的含义。恩格斯首先从哲学角度提出了思维是物质的运动形式的论点。他指出："运动，就最一般的意义来说，就它被理解为存在的方式，被理解为物质的固有属性来说，它包括宇宙中发生的一切变化和过程，从单纯的位置移动起直到思维。"[1]

在现代心理学中，思维被理解为"受社会所制约的，同言语紧密联系的，探索和发现崭新事物的心理过程，是对现实进行分析和综合中间接概括反映现实的过程。思维在实践活动基础上由感性认识产生并远远超出了感性认识的界限。"[2] 也有人认为："思维是人脑对客观现实概括的和间接的反映，它反映的是事物本质与内部规律性。"[3]

在思维科学中，有人把思维看作"是发生在人脑中的

信息变换”。[1] 因为人脑对客观事物的反映不同于镜子的反映。镜子是由玻璃这种物质来反映事物的，人脑是由脑的原生物 —— 意识来反映事物的。而所谓意识，乃是凝聚在脑物质上的信息，即人脑是通过接收与处理客观对象的信息来反映对象的。对象的信息在经过思维后，或者被凝合为意识的信息同化了，或者脑中的意识(信息凝合)被外来的信息改变了，总之是信息的原来形态发生了改变，即信息变换。

尽管不同学科对思维含义的表述各不相同，但细加辨别，其实质是相同的，即思维是理性认识阶段的反映活动。

唯物论认为，认识是主体对客体的反映，这种反映包括感性和理性两个阶段。而思维乃是主体对客体的理性反映形式。因为作为思维器官的人脑，是高度组织起来的特殊物质，它约有 1500 毫升的容量，是黑猩猩脑量的 4 倍，且神经细胞数达 $10^{14} \sim 10^{15}$ 的大脑皮层上还分化出了许多性质不同的区域，故人脑不仅有动物脑那样的功能，能天然地反映客体，产生感觉，而且更具有动物脑所不具备的功能 —— 产生意识，即能自觉摄取知识，积累知识，并凭借知识对事物的本质进行总结、概括、抽象以形成概念，作出判断和进行推理，故它对客体的反映已超越了感性认识的范围，进入到揭示事物本质属性和内部联系的理性认识的阶段。

例如，对刮风、下雨等现象，人和动物一样都能感觉，但将这种现象与吹气、扇扇子、玻璃窗上结水珠、水管子“冒汗”等联系起来，发现它们是“空气对流”的表现形式或“水蒸气遇冷后液化”的结果，这时人们的认识活动已深入到了把握因果关系的思维了。

又如，对图形如梯形的认识，感性认识只能反映各种梯形的形状、大小、颜色，思维则能舍弃梯形的具体形状、大小和颜色等非本质特征，而把其一组对边平行、另一组对边不平行的本质特征概括出来，从而实现了认识从感性到理性的转化。

2. 思维的特性

思维有两个基本特性，即概括性和间接性。

(1) 概括性。思维的概括性首先表现在它能略去同类事物的具体差异，而抽取其共同本质或特征加以反映，这种反映不同感觉和知觉。感觉和知觉都是当前事物作用于人的感觉器官所产生的反映(两者不同的是，感觉是对事物个别属性的反映，而知觉是感觉的综合，是对事物各种属性的整体反映)，是就个别的、当前的事物而言的；而思维却不同，它概括反映的不是个别事物的个别特征，而是一类事物的本质特征。如不管形状、大小、位置、颜色、质地的不同，人们将一组对边平行、一组对边不平行的四边形都概括成一类，称为梯形。可以设想，假如人们不是采用词来标志一般梯形，而采用颜色、大小、形态、位置各不相同的具体梯形来作为研究问题的出发点，那思维的展开将是怎样的艰难。

思维的概括性不只表现在反映客观事物的本质特征上，也表现在它反映事物之间本质的联系和规律上。例如，梯形中的两腰的延长线交于一点，且该点与对角线交点、两底的中点成四点共线；梯形两对角线与两腰组成的两三角形面积相等，且它与两底组成的两三角形的面积比等于两底的平方比等性质，都是揭示事物内部联系的性质，这些性质通过感觉是不能获得的，它只有通过思维，亦即在实践活动的基础上，借助概括、判断、推理后才

能获得。

(2) 间接性。

思维是间接的认识过程。所谓间接，是指认识现在，推测过去，预见未来，都必须以已有的知识、经验或其他事物为中介。思维的间接性也是它与感觉、知觉过程的另一基本区别，即感觉和知觉过程是人对客观事物的直接认识，而思维则是凭借知识、经验对事物进行的间接认识。

例如，早上起身后看到马路潮湿，推测昨天晚上可能下雨，但又见屋顶处干燥，于是断定马路上的潮湿是洒水车洒水的结果。这里，对洒水车在马路上洒过水的认识是通过马路潮湿而屋顶干燥的中介推断出来的，因而是间接的认识。

又如，水分子由 2 个氢原子和 1 个氧原子构成，这凭感觉和知觉是不能获得的。然而，人们凭借已有的知识通过思维把它揭示出来了，这也是间接的认识。

思维的认识的间接性，其原因在于思维是第二信号系统的反映过程。根据巴甫洛夫学说，人的高级神经活动有两种信号系统：一种是以具体刺激物作为信号刺激而建立的暂时神经联系系统，称之为第一信号系统；一种是以词作为信号刺激而建立的暂时神经联系系统，称之为第二信号系统。感觉和知觉是第一信号系统对现实的反映过程，而思维则是借助于以第一信号系统活动为基础，第二信号系统活动为主导的两种信号系统的协同活动。亦即它是以“对前一个系统的无数信号系统的抽象化与概括化的词作为条件刺激物的”。这种词已“超脱了现实”，对它的概括性认识，就是已有的知识和经验，而以这种知识、经验为中介，进行判断、推理，无疑是一种间接的

认识。

思维的间接性与概括性是密切联系的。没有对现实的概括的认识，就无以进行间接的认识。反之，没有对现实的间接的认识，就不能对现实进行更高层次的新的概括。例如，每个词代表着一个概念，每一个概念都概括着一类事物的一般特性，而任何一类事物的数目可以多到无穷无尽，要对这无穷无尽的一类事物进行新的概括，揭示其规律和内部联系，就必须借助于判断、推理等间接的认识过程，故概括性不能离开间接性。但间接认识的基础是作为概括认识的词，故间接性也不能离开概括性。

思维除具有概括性、间接性两个基本特征外，还有自觉性、社会性等特征。

(3) 自觉性。所谓思维的自觉性是指在思维活动中，生理人脑是作为意识的物质承担者而出现的。因为思维的本质是具有意识的人脑对于客观事物的反映。考察纯粹自然状态的人脑，不同人之间差别不大，但思维的效能却有很大的差异，原因就在于大脑中的意识状态的不同。故思维反映客观事物都是程度不同地带着某种自觉性。而感觉不同，许多时候，它对客观事物的反映是自发地进行的。

(4) 社会性。所谓思维的社会性是指思维是社会实践与历史发展的产物。因思维超越了感性认识的范围，深入到事物的内部，从而可以揭示规律，预言未来，故思维对于存在具有相对独立性。同时，由于思维常以理论形式出现，其形成的结论是否正确，必须加以检验。这种检验，不管采用什么方式，最终的标准仍是实践。思维的社会性还指思维必须以已有的知识为前提，而这些知识正是社会历史发展的产物。

1.2 数 学 思 维

1. 数学思维的特性

应该怎样认识数学思维?曾有一位数学家作过如下的论断:“数学是一种思维形式,它牢固地扎根于人类智慧之中……数学表现了人类思维的本质和特征,并在任何国家与民族的文明中都会有所体现,因而在当今的意义下,任何一种完善的形式化思维,都不能忽视这种数学思维形式。”[4] 这段话的意思无非是指数学是思维的一种形式,因而它包含一般思维所具有的本质,即数学思维是理性的认识活动,同时它还表现出数学学科本身的特殊性。综合两个方面,数学思维的特性一般地表现为:

(1) 数学思维较之其他思维具有更强的间接性和概括性。由于数学高度抽象的特点,使得数学思维较之其他思维更为间接,亦即是间接的间接。例如,数学概念是对数学本质的抽象,这种抽象只保留了事物量的关系和空间形式,而舍弃了其他自然性质,这种抽象程度是其他学科所不及的,故任何思维都没有数学思维来得间接。与此相联系的是,数学思维的概括是概括的概括。因为不仅数学概念是抽象概括的产物,而且数学推理、方法等也是各种具体经验、方法的升华,数学思维是在这些基础上进行的概括,故它是概括的概括。

(2) 数学思维具有独特的形式化的符号语言。数学思维的这一特性,正如 M. 克莱因(M. Klein) 所指出的:“数学的另一个重要特征是它的符号语言。如同音乐利用符号来代表和传播声音一样,数学也用符号表示数量关系和空间形式。” 他又说:“数学语言则是慎重地、有意识

地而且经常是精心设计的，凭借数学语言的严密性和简洁性，数学家们就可以表达和研究数学思想，这些思想如果用普通语言表达出来，就会显得冗长不堪。这种简洁性有助于思维的效率。”[5] 这表明，数学思维的符号化特性是与这种思维高度的间接性、概括性相适应的。因为只有运用摒弃了具体内容的形式化符号，才能保证思维实现更高程度的抽象和概括。

(3) 数学思维具有显著的美学特性。美学特征被概括为统一性、简洁性和奇异性三个方面。数学思维由于其高度的间接性和概括性，保证了数学思维的统一性和奇异性。因为只有高度抽象、高度概括才能达到高度统一，而高度间接性的思维反作用于现实，就隐含了思维的奇异性。同时，数学思维的符号性特征又保证了思维的简洁性，因此数学思维往往具有显著的美学特征。数学思维的这种美学特征，正如庞加莱(H. Poincaré) 所指出的：“我敢冒昧地说，数学的探索还有深刻的美学原则。”[4] 又如A. 波莱尔(A. Borel) 所说的：“我们的活动与艺术家的活动有许多共同之处：画家进行色彩与形态的组合，音乐家把乐音组合起来，诗人组词，而我们则是把一定类型的概念组合起来。”[5] 可见，数学思维的美学特征是十分明显的。

(4) 数学思维具有独特的辩证性。数学中充满着各种形式的辩证关系：形与数、量与质、常与变、曲与直、有限与无限、近似与精确、离散与连续、抽象与具体、特殊与一般、正运算与逆运算等的相互对立、相互制约、相互联结、相互转化的关系，在数学中表现得最为充分。辩证法的三大规律在数学中也得到了很好的体现。据此，可以这样说，数学思维在一定意义上有时是指这样的一种形式，

它表现为人们认识具体的数学科学，或是将数学应用于其他科学、技术、国民经济等的过程中的辩证思维。正如恩格斯所指出的："数学：辩证的辅助工具和表现方式。"[6]

2. 数学思维的品质

思维的发生和发展，既服从于一般的、普通的规律性，又表现出个性的差异，这种个性差异表现在个体思维活动中的智力特征方面，就是思维品质。数学的思维品质是由数学思维本身的特征所决定的，其考察指标主要有如下几方面。

(1) 思维的深刻性。数学思维的高度概括性、间接性决定了思维的深刻性是衡量数学思维品质的一个主要标志。所谓思维的深刻性，是指分清实质的能力。这种能力表现为能洞察所研究事物的本质及其相互联系；能从所研究的材料中揭示被掩盖的特殊情况；能组合各种具体模式等。

例如，已知矢量等式

$$\vec{a}^2 = (\vec{b} + \vec{c})^2, \qquad (*)$$

应明确其实质就是余弦定理。但由于矢量等式具有比余弦定理更简洁的表达形式，且利于应用某些代数运算性质，故用矢量等式(*)来证明某些空间关系问题往往更新颖，更简洁。比如：

"四面体 $ABCD$ 中，$AB \perp CD$，$AC \perp BD$，则

$AB^2 + CD^2 = AC^2 + BD^2$。"

若设 A、B、C、D 对空间某一点 O 的矢径为 $\vec{r}_1$、$\vec{r}_2$、$\vec{r}_3$、$\vec{r}_4$，则由条件知

$$(\vec{r}_2 - \vec{r}_1)(\vec{r}_4 - \vec{r}_3) = 0,\ (\vec{r}_3 - \vec{r}_1)(\vec{r}_4 - \vec{r}_2) = 0。$$

两式展开整理后，得

$$\vec{r}_1 \cdot \vec{r}_3 + \vec{r}_2 \cdot \vec{r}_4 = \vec{r}_1 \cdot \vec{r}_2 + \vec{r}_3 \cdot \vec{r}_4,$$

从而有

$$(\vec{r}_1 - \vec{r}_3)^2 + (\vec{r}_2 - \vec{r}_4)^2 = (\vec{r}_1 - \vec{r}_2)^2 + (\vec{r}_3 - \vec{r}_4)^2,$$

即　$AC^2 + BD^2 = AB^2 + CD^2$。

(2) 思维的广阔性。数学思维的高度的概括性和间接性也决定了它的应用的广泛性，故思维的广阔性也必然成为衡量数学思维品质的一个重要标志。所谓思维的广阔性，即是指思路宽广，善于多角度、多层次、多方位地思考问题，不仅能研究问题的本身，而且能由此及彼地思考与其相关的其他问题等。

例如，完成一个不等式的证明，按思维的广阔性要求，不仅能就给出的不等式直接加以证明，而且还要能广泛地联想，通过各种变形，构造辅助函数等手段加以解决。比如：

“设 $x > -1$，且 $x \neq 0$，n 是不小于 2 的整数。证明：$(1+x)^n > 1 + nx$。”

在具体思考时，不仅应考虑到它是与自然数 n 有关的命题，可对 n 用数学归纳法加以证明，还应考虑到通过变形，原不等式可化为

$$(1+x)^n - 1 > nx,$$

从而问题又可转化为证：当 $x > 0$ 时，有

$$\frac{(1+x)^n - 1}{x} > n;$$

而当 $-1 < x < 0$ 时，有

$$\frac{(1+x)^n - 1}{x} < n。$$

更应考虑 $\frac{(1+x)^n - 1}{x}$ 恰是函数 $f(t) = (1+t)^n$ 关于

$\dfrac{f(x)-f(0)}{x-0}$ 的表达式，其形状像拉格朗日中值定理，从而想到用拉格朗日中值定理证明。又考虑

$$\frac{(1+x)^n-1}{n}=\int_0^x(1+t)^{n-1}dt,$$

故原不等式可化为证

$$\int_0^x(1+t)^{n-1}dt>\int_0^x dt。$$

再者，还应想到，如果令

$$f(x)=(1+x)^n-nx,$$

则 $1=f(0)$。这样，问题变为证 $f(x)>f(0)$，即可归结为 $f(x)$ 的最值问题，等等。

(3) 思维的灵活性。思维的灵活性是指能根据条件的发展变化，及时改变先前思维过程，寻找解决问题的新途径。亦即思维不局限于固定的程式，还能及时摆脱心理定势的影响。

例如，对于"在 $\triangle ABC$ 中，求证：$\cos^2A+\cos^2B+\cos^2C=1-2\cos A\cos B\cos C$"的问题，按照程式化的思考方式应归结为应用三角公式和余弦定理证明，但若考虑将右端移至左端，且将 $2\cos A\cos B\cos C$ 拆成两个 $\cos A\cos B\cos C$，则左端便是如下的三阶行列式：

$$\Delta=\begin{vmatrix}-1 & \cos C & \cos B\\ \cos C & -1 & \cos A\\ \cos B & \cos A & -1\end{vmatrix}$$

于是，问题变为通过行列式变形化为证 $\Delta=0$。注意到若在第一、第二、第三列上分别用 a、b、c 去乘，然后把第二、三列都加到第一列上，应用射影定理，便得

$$\Delta = \frac{1}{a}\begin{vmatrix} 0 & \cos C & \cos B \\ 0 & -1 & \cos A \\ 0 & \cos A & -1 \end{vmatrix} = 0,$$

这样的思考就比直接用三角公式的思考要灵活得多。

(4) 思维的创造性。在心理学中,"凡是能提供新的、独创的、有高度社会价值的活动",被称为"创造活动"[7]。这里,思维的创造性是指独立思考出有社会(或个人)价值的具有新颖性成分的成果的智力品质。它的特点是主体对知识、经验和思维材料进行新颖的综合分析、抽象概括,以致达到人类思维的高级形态,其思维结果包含着新的因素。

例如,一般人的头脑中只有唯一的现实空间,而数学家们创造了四维空间,五维空间,…,n 维空间,非欧几何空间,拓扑空间,超限数空间等;正方形与圆既不全等,又不相似,但从拓扑观点看,它们的结构是相同的,如它们都将一个平面划分为两部分,其内部区域的连续性是一样的,在拓扑变换下,两者可以互变。像这样的思维就具有推陈出新、常中见异的特点,因而具有创造性。

然而创造性是有层次的,对于数学的学习者来说,其思维的创造性通常表现为分析、思考、理解数学知识和解决数学问题中的某种独特见解和独创解法。

例如,给出式子"$7 > 2\pi$",围绕此式展开联想。有人说,由给定式子可得出:若把地球赤道看成一圆,假想有一根绳子绕地球赤道一圈,正好使绳和地球赤道等长。这时,若把此绳延长 7 米,在赤道外面围一个同心圆,则这个圆与赤道各点的距离要超过 1 米。显然这种联想是基于不等式 $2\pi r + 7 > 2\pi(r+1)$ 展开的。无疑,这样的思维对于学习者来说,也是一种创造性的思维。

(5) 思维的敏捷性。思维的敏捷性即是指思维能通过高度压缩的形式在很短时间内对事物作出迅速的判断。思维的敏捷性与记忆密切相关。只有实行有条理的记忆,并通过各种记忆手段强化记忆功能,才可能在思维过程中实现经济原则,达到思维的敏捷性。离开了记忆,思维的敏捷品质是无从谈起的。

检验数学思维的品质,除以上诸方面外,还有思维的目的性和批判性。前者是指思维的方向要围绕着总目的,并通过不断探索与选择,逐步寻求达到总目的的途径,进而实现总目标;后者是指思维活动中善于严格地估计思维材料和精确地检查思维过程的品质,它是思维过程自我意识作用的结果。对于学习者来说,主要是指对教材、参考书、教师所传授的知识和自己的解题作业的严格检验等。

3. 数学思维的成分

数学思维就其基本成分而言,有具体形象思维、抽象逻辑思维和直觉思维三种,它们分别属于三种不同层次的思维。

(1) 具体形象思维。形象思维是建立和理解数学概念的基础。在数学科学的演绎体系中,从原始基本概念到命题、结论,无一不是从具体事物中抽象出来的理想事物。显然,对理想事物的形象感知,纯逻辑、纯抽象并不能够解决全部问题,而具体形象思维则在其中起着应有的作用。

以图表为例。在数学中,图表不仅是研究的对象,而且也是数学的重要语言。用图表表达数学内容,考察对象就形象地呈现在人们面前,且易被人们的心灵感知、记忆和应用。

应指出的是，数学中的形象思维已不是形象思维低级阶段的形态，亦即不是单凭人的感官在所能感知的范围内的形象思维，而是在前一步抽象的基础上创造的形象思维。它包括将客观事物中的数学特点抽象出来而构成的模型、表格、图形等，也包括数学知识在头脑中形成的记忆表象。数学中的形象思维是图表、言语等直观教学得以顺利进行的根本条件。

(2) 抽象逻辑思维。数学中的抽象逻辑思维有形式逻辑思维和非形式逻辑思维。数学思维之抽象性、概括性和辩证性在这里表现得最为充分。限于篇幅，我们不加讨论。这里特别指出以下两点。

第一，数学中的抽象对于数学的抽象逻辑思维具有十分重要的意义。例如，形式逻辑思维包括概念、判断、推理、证明等，无论是哪一种思维，其过程都离不开抽象这一思维形式，其中所谓弱抽象和强抽象对于由已知概念去引出新概念是至关重要的。弱抽象是指由原型(或已有概念)中选取某一特征(侧面)加以抽象，从而获得比原结构更广的结构，也即使原结构成为后者的特例。例如，从函数的实例中抽象出函数概念是弱抽象，这种抽象层次越高，概念脱离现实原型就越远，而反映事物类的面却越广。这相当于逻辑学中所说的“外延广，内涵少”。而强抽象，是指通过引入新特征强化原结构来完成的抽象，其所获得的新结构是原结构的特例。如在映射的基础上，通过引入“单”和“满”这些新特征得到“单射”和“满射”的概念，便是特征强化抽象，亦即强抽象。这样的抽象层次越高，切近事物本质就越深，而其反映面却越窄。

第二，辩证逻辑思维是一种非形式逻辑的思维，目前对其表现形式及其规律的探讨还很不充分，有待进一步

研究和发掘。不少学者认为："辩证逻辑该是一般辩证法在思维领域的特殊表现。""辩证逻辑，对于思维科学而言，是高级的规律。辩证逻辑并不取消、代替任何科学的逻辑，而是运用与提高各门科学的逻辑。这和高等数学并不取消初等数学、爱因斯坦物理并不取消牛顿物理、现代几何并不取消欧几里得几何学的道理，是毫无二致的。"（孙叔平为李廉著《辩证逻辑》的序）数学中的辩证逻辑思维的表现形式虽有很多，但最基本的是以形式逻辑思维为基础，在对立统一规律的指导下，溶解形式逻辑固定分明的界限，使认识与客观世界相吻合。这样的辩证因素在概念形成、猜想获得和规律发现的过程中都有着充分的展示。

(3) 直觉思维。对直觉思维赖以进行的心理过程和生理结构的研究表明，直觉思维与下面两种无意识的心理现象有联系，即大脑中潜存的信息及大脑两半球的交互作用。直觉思维的形式主要是想象和判断。它或者是创造性想象接通了已知和未知的通路，从而形成整体形象，并把握事物本质作出判断；或者是抓住其中一两个具有本质特征的东西核实后进行判断。因此，这样的判断不同于逻辑判断，是整体判断或特征判断。关于数学中直觉思维的具体论述，已有不少著作论及，有兴趣的读者可参阅有关书籍。

数学思维就其基本成分而言有上述三种。这三种思维成分贯穿于一切思维过程的始终，并且它们相互联结，相互作用。例如，直觉思维本质上乃是一种最高层次的形象思维。又，应该区分作为思想成分的具体形象思维、抽象逻辑思维、直觉思维，与所谓辐射、辐集思维，正向、逆向思维，相似思维，分合思维的不同。辐射思维是"从同一

的来源中产生各式各样为数众多的输出”的思维；辐集思维是以某个思考对象为中心，从不同方向将思维指向这个中心的思维；正向、逆向思维是就思考数学问题的方向而言的；相似思维是对不能建立函数关系的两事物的相似性的考察与类比，而分合思维则是思维过程中的分析与综合。实际上它们都是以不同的标准来研究思维过程的产物，它与思维的三种基本成分毫不相同，不能混为一谈。

4. 数学思维的结构

数学思维过程构成了一个包括数学知识、方法及主客体交互作用的系统。系统论的整体原则表明，任何系统都是有结构的，且系统整体的功能大于各孤立部分的功能之和。数学思维结构是思维结构的重要的子系统。研究数学思维结构可使我们从整体上把握数学思维系统，明确数学思维过程中的各种因素、各因素间的相互作用及整体规律等。在动态中协调内外各种因素，使所有部分的功能都能服从于整体的最优目标，从而取得比较理想的整体效益。

瑞士心理学家皮亚杰认为，思维结构和数学科学结构十分相似。因此，数学思维结构是思维结构和数学知识结构的一种特殊综合。

一般认为，数学思维结构是主体在思维中以数学知识、理论为基础在头脑中建立起来的信息操作系统。其结构是一个三维立体，三轴之一是数学思维的基本成分，其内容已在上节阐述。另外两轴分别是个体发展水平和数学思维方式。

(1) 个体发展水平。数学思维的发生和发展既有一般意义的普遍规律，又表现出个体差异。表征这种差异的

是思维的智力品质和非智力品质。

数学思维的智力品质已在本章论述。它反映了人们在进行数学思维时,无论是形象的、抽象的、辩证的,其操作在深度、广度、速度、灵活程度、批判程度等方面的不同情况,体现了思维的层次性,它是度量人们思维是超常、正常还是低常的一个重要指标。

表征个体发展水平的另一特征是思维的非智力品质,这包括动机、情感、意志等内容。

在数学思维活动中,动机是引起个体行为的内在动力,其作用是促使人们进行有目的的行动。在思维过程中,动机是通过加强努力、集中注意、积极活动而促进思维活动的。此外,积极的态度,顽强的毅力是进行繁难、艰苦的数学思维活动所不可缺少的。总之,非智力因素是数学思维内驱力的巨大源泉,它从根本上决定着一个人能否进行正常有效的数学思维活动,对于青少年学生来说,尤为如此。

(2) 数学思维方式。对于思维方式,列宁曾有过一个精辟的定义。他说:人的认识活动,客观上存在着三个要素:认识主体(= 人脑);认识对象(= 自然界);认识工具(= 思维方式)[8]。亦即思维方式是连接主体与客体的桥梁,是实现一定认识目的的工具。数学思维方式是人们在学习数学知识,理解数学观点、数学精神、数学思想,掌握和应用数学知识、方法的基础上形成的。其中,数学知识、数学观点、数学精神、数学思想和数学方法等要素发挥着重要作用。知识要素包括理论与经验,是思维方式赖以产生、形成的基础与必要前提,它是思维方式最基础的层次;观念、精神、思想要素是在一定的知识结构基础上凝结的对于事物的观点、意念的总和;方法要素是人们思考

处理问题的方式和方法，这种方式、方法也是一个多层次的结构系统，包括一般方法、特殊方法和具体方法等。

由此可知，数学思维方式作为主体进行数学思维活动的形式，是数学知识、数学观点、数学精神、数学思想、数学方法的多级系统的综合体，又是数学思维结构的重要组成部分，无论是思维的深度、广度、速度、创新程度等方面，思维结构都受到思维方式的影响。反过来，任何客观的思维方式都不是同个体无关的，它必然存在于个体思维之中，通过个体的思维结构发挥作用。

上面我们分析了数学思维结构的三个维度及其它们的相互关系。可以看出，数学思维结构形成的速度和完善程度，不仅与学生的生理素质有关，而且与教育有密切的关系。故如何根据人类认识的总规律，根据数学科学本身的特点和学生个体认知的特点进行教育，以完善和发展学生的思维结构，是数学教育、教学的首要任务。

日本数学家、数学教育家米山国藏曾说："我搞了多年的数学教育，发现学生们在初中、高中等接受的数学知识，因毕业进入社会后几乎没有什么机会应用这种作为知识的数学，所以通常是出校门后不到一两年，很快就忘掉了。然而，不管他们从事什么业务工作，唯有深深地铭刻于头脑中的数学精神、数学的思维方法、研究方法、推理方法和着眼点等（若培养了这方面的素质的话），都随时随地发生作用，使他们受益终生。"[9] 米山国藏的精辟论述，深刻揭示了作为数学思维方式的数学观点、数学精神、数学思想、研究方法、推理方法、着眼点等在人们思维活动中的重要意义。

二 数学的基本思维方式

著名数学家韦尔(H. Weyl)在《数学的思维方式》的报告中指出:"所谓数学的思维方式,首先是指数学用以渗透到研究外部世界的科学(例如物理学、化学、生物学和经济学等),以及渗透到我们人类事务的日常思维活动中的那种推理形式,其次是数学家们应用于自己的研究领域中的推理形式。"[10] 本章所展示的各种数学思维方式,无论对于研究外部世界的科学,还是进行日常思维活动和数学的自身研究,都具有普遍的指导意义,故称之为数学的基本思维方式。

如同任何事物都是一种历史的产物一样,数学思维方式也是人们在认识世界、改造世界的实践中形成和发展起来的。谈论数学思维方式,离开其产生和发展的历史是不客观的。故本章在论述各种思维方式时,除了阐明其基本原理、思想、功能及其应用外,我们还将力求反映其真实的历史侧面,尤其反映蕴涵在中国古算中思维方式的丰富思想,因为这在以前常被人们所忽视。

2.1 符号思维方式

符号思维方式是数学中最为基本的思维方式,它是通过设计符号、运用符号来进行分析、思考和推理论证,从而去实现数学中的创造、发明。

1. 符号、数学符号及其特征

所谓符号，是人们约定用来指称一定对象的标志物，是用以表达和交换思想的工具。符号最基本、最重要的形式是语言符号，而语言符号又可分为自然的和人工的。所谓自然语言符号，是指每个时代历史地形成和存在的生动的民族语言的总和，一般它是"民族语言"的同义语。由于自然语言存在着多义(歧意)性，语法的复杂性和结构的非模式性等，对科学的表述存在一定的困难。因此，数学和各门科学为了适应自身的发展，就需要专门建立一种符号体系，这就是人工语言符号。这种人工语言符号，是由基本符号和合式公式形成的规则两种成分组成的，这是一种完全形式化的语言。"人工语言"加上"推理规则和公理"便是形式系统。通常的"数学语言"乃是自然语言加上一批数学符号构成的一种语言，它是数学共同体专门约定的一种语言。

数学符号，作为符号的一种，它首先具备一般符号的共同性。

(1) 物质性。作为对象标志物的符号首先是以自己的物理表现给人以感性映象的。一切符号，在形式上都表现为一定的物质运动或物质存在形式，且经过了人的加工或改变，带有人的痕迹的物质形式。数学符号是数学共同体，通过实践的力量创造出的自然界原来并不存在的一种新的物质形式。这是一种用感官可以感到的物质形式。

(2) 非相似性。符号与模型、映象不同，它与所标志的对象之间的关系是假定的、任意的、非相似的。对此列宁曾指出："如果感觉不是物的映象，而只是同物没有'任何相似之处'的记号或符号，那么作为赫尔姆霍茨的出

发点的唯物主义前提就被推翻了，外部对象的存在就有些问题了，因为记号或符号完全可能代表虚构的对象。”[11] 例如，被称为哈密尔顿算子的倒三角形“▽”，表示：把它用于数量函数 $u(x,y,z)$ 得一向量

$$\nabla u = \frac{\partial u}{\partial x}\vec{i} + \frac{\partial u}{\partial y}\vec{j} + \frac{\partial u}{\partial z}\vec{k},$$

称为 u 的梯度，表示 u 在点 (x,y,z) 一切方向的变化率中最大者的方向。其符号的倒三角形形状与代表的意义无任何相似之处。当然，并不排斥某些数学符号与所表征的某些事物具有一定的相似性，如三角形符号“△”，平行四边形符号“▱”等。

数学符号作为数学科学的人工符号，它又具备区别于其他符号的一些不同特征。

(3) 抽象性。任何符号，分析其构成，都是一种联系能指与所指的两面体。数学符号区别于其他符号的一个特点是，它所指的对象具有高度的抽象性。因为数学所研究的已是一种形式化的思想材料，或者如布尔巴基(Bourbaki) 学派所说的，是一种“研究抽象结构”的理论。例如，世界上本没有方程，方程是人们从现实世界量关系中抽象出来的思想材料。与此相对照，没有人固然没有化学，但元素碳、氢、氧还是客观存在于人脑之外的现实中。因此，数学符号指代的对象，抽象程度要高于其他科学符号，且在抽象程度上具有不同的层次性，不仅有直接概括对象属性的抽象，还有在此基础上的更进一步的抽象。例如，具体自然数符号 1，2，…，一般自然数符号 n，自然数集合符号 N 等就是不同抽象程度的数学符号。

(4) 精确性。数学符号作为本质上是一门推理科学的人工语言符号，它具有精确性。因为数学的对象是形式

化的思想材料，它的结论是否正确，一般不像物理学那样，通过重复实验来检验，而主要地要靠严格的推理来证明。如果符号不精确，就很难保证推理的正常进行。在微积分发展史上，牛顿曾用"0"代表无穷小量，但这无穷小量是什么，是"固定量零"，还是"趋于零的量的瞬"，牛顿也说不清楚。他时而这样说，时而又那样说，这就导致了数学史上的第二次危机。它从一个侧面证明了数学符号的精确性对数学科学的重要意义。

(5) 规范性。符号与它指代的对象必须相对稳定，否则，任意改变符号的意义或无规则地乱用符号，就会使符号失去信息传递的作用，故数学符号能指与所指的联系，一经规定后就约定俗成具有规范性。

(6) 通用性。在现代科学中，只有数学符号系统是统一的，不论哪个国家，也不论哪个民族的数学家，他可以很方便地交流数学成果。从最简单的对象符号 1、2、3、4、5，到最复杂的组合符号数学公式，尽管各国的读音不相同，但其意义是相通的。

从数学符号整个系统考察，它又具备以下特征。

(7) 自我生成性。数学可以构成如下的符号系统，它能把整个理论用纯符号的形式表达出来。这样，只要符合规则，我们就可以从已有的符号中产生新的符号。其中复数符号 $a+bi$ 就是一例。16 世纪，被誉为"第一流的代数学家，严肃的哲学家，具有精确观测习惯的物理学家和技艺高超的医生"的卡尔丹诺(Cardano)，一次在考虑"怎样将 10 分成两部分，使两者的乘积等于 40"时发现，如果把 10 分成 $5+\sqrt{-15}$ 和 $5-\sqrt{-15}$，那么不管它们代表什么，结果都是对的。于是就产生了形如 $a+b\sqrt{-1}$ 或 $a+bi$ 的记号。故复数记号实际是记号 a、

b 及单位 1、$\sqrt{-1}$ 的生成物。

(8) 开放性。数学符号系统是一个开放系统，它是随社会实践的发展不断容纳新符号而使自己逐步完善的。例如，随着科学技术的发展，人们要处理和研究的对象越来越复杂，由互克性原理，“当系统的复杂性日益增长时，我们作出系统特性的精密而又有意义地描述的能力将相应降低”，这样，模糊量方法于 20 世纪六七十年代就应运而生，而模糊量符号系统也就被纳入到数学符号系统的大家庭中了。

2. 符号方法的历史演变

纵观至目前为止的符号方法的历史发展，可以大致区分为萌芽时期、产生时期和形式化时期三个阶段。从古代到 16 世纪符号代数出现之前是萌芽时期，从 16 世纪到 19 世纪末为产生时期，从 19 世纪末起，数学符号开始进入形式化时期。

(1) 萌芽时期。

从古代到 16 世纪前，人们对数学符号的认识还是朦胧的。远在有文字记载之前，数的概念和计算方法，以及对各种形状的认识就已经发展起来了。随着生产实践的发展，伴随着计数，许多文明发达较早的民族逐渐产生了各自的记数符号和记数方法，经过长期的、一定范围内的交流、传播和比较、鉴别，加之 1450 年，约翰·古登堡活版印刷的发明对数码规范化的促进，欧洲到 16 世纪或以后更长些时间，逐渐形成了目前比较通行的阿拉伯记数符号和记数方法。在代数和几何领域中，只是不成系统地出现了一些缩写和象形符号。

人们记数，经历了从原始符号记数至数字符号记数的发展阶段。其中，原始符号记数包括实物记数、结绳记

数和刻痕记数等。

约在距今七八千年前的新石器时代，人类开始了定居，并出现了一定量的农牧业，劳动的发展和交往的增多促进了记数的产生。开始，数符号是用实物指代的。例如，至今还有些原始部落的人仍用身体器官指代数量。他们顺次从左手小指、无名指直到拇指分别指代 1、2、…、5，用腕、手肘之间直至颈分别指代 6、7、…、10，等等。后来，由于实物指代不能满足需要，便产生了结绳记数和刻痕记数。正如我国《易传・系辞》中记载的，“上古结绳而治，后世圣人易之以书契”。[12] 这里，“实物”、“结绳”和“刻痕”都属原始的记数符号。

原始记数符号具有很大的局限性，它的进一步发展就产生了数字记数符号。世界文明发达较早的国家都产生了各自的记数符号。

运算符号并非随着运算的产生立即出现。古埃及、中国、希腊和印度长期都没有加法符号，把两个数字符号写在一起就表示相加，在今天的分数写法中，我们仍可以看到这种痕迹。直至 3 世纪时，希腊才出现减号“↑”；6 世纪，印度出现缩写符号，其中乘法用 bhavita（乘）的缩写 bha 表示，减法则在减数上画出一点来表示，加法和除法仍无符号。

对符号的注意是从欧洲开始的。起先，欧洲人承袭印度的做法，用单词的第一个字母作符号。如用拉丁字母的 P(plus 的第一字母，意为相加）表示加，用 M(Minus 的第一个字母，意为相减）表示减。后来，到了 16 世纪才形成了近似于目前使用的精细的运算符号。

代数在这一时期基本上都是文辞的和缩写的。与巴比伦和埃及的一样，古希腊代数基本上是文辞的，只是在

3 世纪，丢番都(Diophantus) 有意识地应用了缩写符号后才进入缩写阶段。印度的代数是缩写的，且缩写的程度强于丢番都。例如，婆罗摩及多(Brahmagupta，约 628) 及其后人，就用 ca 1644ni lru 6302 表示 $1644y+z+6302$。其中，ca 是 calaca(黑的) 缩写，用以表示第二未知数。ni 和 ru 分别是 nilaca、rupa 的缩写，表示第三未知数和单纯数，且几部分并列在一起表示相加，没有加号。

几何领域中，在古埃及的数学纸草和巴比伦的数学泥版中，已发现在图形上标有尺寸，但还没有出现标注字母的情况。公元 1 世纪，海伦(Heron) 在《经纬仪》一书中，曾用 △ 表示三角形，$\underline{\underline{ov}}$ 或 $\underline{\underline{p'}}$ 表示平行和平行四边形。3 世纪，帕普斯(Pappus) 用 ○ 和 ⊙ 表圆，∠ 表角，□ 表正方形。1202 年，意大利斐波那契(Fibonacci) 开始用 a、b、c、… 等小写字母表示点等。但这些符号还都是零星的、象形的，且带有极大的随意性。后来，在代数符号化的推动下，加之 17 世纪代数与几何结合产生解析几何，促进了几何图形符号的急速发展。到 19 世纪，法国几何学家卡诺(Carnot) 的著作《论几何图形的关系》(1801) 中，几何图形的符号已很接近于现在了。如 A、B、C 表示点，$\overline{AB}$、$\widehat{AB}$ 分别表示线段 AB 和弧 AB，$\overline{BCD}$ 表示 C 在 B、D 之间的共线三点 B、C、D，$\overline{AB}\cdot\overline{CD}$ 表示直线 AB 与 CD 的交点，等等。

(2) 产生时期。

从 15 世纪起，欧洲爆发了一场称之为文艺复兴的深刻的资产阶级民主革命，前后历时 200 多年。新兴资产阶级反封建的革命斗争，为科学发展扫清了道路。而科学的发展又要求作为启发思维有效手段的数学创造一种更适

合于它发展的符号语言。加之此时，我国造纸、印刷术传入欧洲后欧洲发明了活版印刷，更为数学符号的创用提供了物质条件。这一时期，人们对数学符号的认识已从朦胧阶段进入到有意识的创用阶段。这突出地表现在代数符号的变革和变量数学符号的产生上。

代数史上，数学符号的创用使代数从文辞的、缩写的进入到了符号的阶段。这一转变的主要标志有三方面：一是表示未知数和已知数符号的出现；二是合理的幂次符号的建立；三是精巧的运算符号和关系符号的产生。

未知量、已知量符号的出现是代数走向符号化的重要一环。因为没有它，数学便不能从特殊走向一般。16 世纪的法国数学家韦达(Vieta) 是数学史上第一个有意识地系统使用字母的人。1585 年，他创用大写元音字母 A、E、I、O 等表示未知数，辅音字母 B、G、D 等表示已知数，并称使用字母后的代数方程所表示的是“类的计算术”，以别于“数的计算术”，以此作为代数与算术的分界线。但他的符号体系还是混乱的，言词常与缩写混在一起，没有等号，乘号用“in” 表示。在这后，笛卡儿(Descartes) 对韦达的字母作了改进，他用字母中前面的字母表示已知量，末后的表示未知量，成为现今的习惯用法。

幂次符号是伴随着代数摆脱几何束缚趋势的出现而产生的。起先，人们往往借助平方、立方来表示幂。例如，丢番都用 $\triangle^y$、k^y 分别表示 x^2 和 x^3，而用 $\triangle y\triangle$、$\triangle k^y$ 分别表示 x^4 和 x^5，随着代数摆脱几何束缚趋势的出现，人们强烈指责滥用“平方”、“立方” 的做法，主张应直接使用“四次”、“五次” 等的用语。居住荷兰的机械师、工程师斯台文(Stevin) 首先创用了符号 ◯ 表示未知量，并将次数写于圈内。后来，笛卡儿对此作了改进，他用

$$x^{iv} + 4x^{iii} - 19x^{ii} - 106x - 120$$

表示

$$x^4 + 4x^3 - 19x^2 - 106x - 120,$$

从而首先给出了正整指数幂的现代表示法。在笛卡儿工作的基础上，沃利斯(J. Wallis)、牛顿(Newton)还创用了零指数、负指数和分数指数幂记号。至此，统一分式、根式的科学的幂次符号便得到了完全的确立。

数学符号按其性质可分为元素符号、关系符号、运算符号、约定和辅助符号，代数从文辞的、缩写的进入到符号的，除有元素符号外还必须有运算符号、关系符号甚至还要某些约定和辅助的符号，因为它们都是简化逻辑推理所不可缺少的。

近代精细的运算符号和关系符号是在缩写符号的基础上发展起来的。"+"、"-"号开始时是作为"超过"和"不足"由德国数学家维特曼(Widmann)于1489年引进的，直到1514年，荷兰赫克(Hoecke)才赋予它加、减号的意义。乘号"×"是英国数学家奥特里德(W. Oughtred)于1631年在《数学的关键》一书中首先使用的。1698年，莱布尼兹(Leibniz)又创用了点乘号"·"。除号"÷"是瑞士数学家拉恩(Rahn)于1659年首先使用并由牛顿、沃利斯提倡的。据说，它是由曾经用作过除号的"—"与"："的组合。关系符号"="是英国列科尔德(Recorde)于1557年引入并由莱布尼兹倡导的。列氏说："最相像的两件东西是两条平行线，所以这两条线应该用来表示相等。"辅助符号圆括号始见于1544年，方括号是沃利斯首先使用的，花括号是韦达于1593年引入的。正是由于这样三个方面符号的出现，代数在十六七世纪才从文辞的、缩写的进入到符号化的阶段。

被誉为"人类精神的最高胜利"的微积分，其符号是由它的理论创立者牛顿、莱布尼兹各自独立创造的。牛顿把变量叫做流，变量的变化率叫做流数。对于流 x、y 的流数记以 $\dot{x}$、$\dot{y}$，其记号首先见于他 1736 年《流数法和无穷级数》一书。莱布尼兹于 1684 年在题为《关于极大和极小以及切线的一种新方法，它对分数或无理数也适用》的文章中，首先使用微分记号 d。积分记号"∫"是莱布尼兹与雅各·贝努里(Jacob Bernoulli)在通信讨论中共同创造的，∫是"和"sum 一词第一字母的拉长。微积分符号的一次重要飞跃是"ε-δ"符号语言的产生，它是柯西(Cauchy)、外尔斯特拉斯(Weierstrass)为克服模糊不清的无穷小量所带来的麻烦而建立起来的。这种符号语言首创了完全不依赖于几何直观而对极限、连续等一系列重要概念的静态刻画，从而结束了微积分发展史上的种种混乱局面，有力地推动了整个分析数学的发展。

(3) 形式化时期。

19 世纪末，数学在刚刚解决了由微积分概念的模糊性而带来的种种麻烦后不久，又出现了集合悖论，这更进一步促进了人们对数学基础的研究。这种研究推动了数学符号向更高方向的发展。

首先，人们从逻辑和语言上对数学基础进行研究，引进了量词符号。这首见于弗雷格(F. L. G. Frege)《表意符号》(1879)一书。在此书中，他完备地发展了命题演算，又几乎很完备地发展了谓词演算。但弗雷格的符号和历来相传的、当时使用的、迄今使用的都完全不同，以致他的学说长期没人注意。1885 年，皮尔斯(Peirce)引进了当今仍使用的存在量词、全称量词符号 Σ_λ 和 π_λ。量词的引入对简化数学表达和推理具有重要意义。因为在近代数

学中，到处充满有关量词的语句和推导。例如，作为微积分基础的极限概念的定义中就有三个量词重叠的语句。自弗雷格以后，意大利数学家皮亚诺(Peano) 于 1894 年出版了《数学公式》。他正式利用前人命题演算和谓词演算的成果，用以表述数学，推导数学。至今所沿用的记号大多是由皮亚诺订定的。此外，他又区分了集合论中 $\in$（属于）和 $\subset$（包含）符号。在皮亚诺稍后些，罗素(Russell) 和怀特海(Whitehead) 系统总结了普通逻辑符号化、数学化的研究成果于《数学原理》一书，成为数学逻辑成熟的主要标志。

数学符号形式化的另一标志是公理化方法从抽象到形式化的发展。这是希尔伯特(Hilbert) 在数学基础研究中，为了实现希尔伯特规划而采取的重要步骤。所谓形式化，用他自己的话说，就是“取代了那种用自然语言来表达的、具有内容的数学知识，我们最后得到的是由数学符号和逻辑符号所组成的一组公式。这些公式是按照一定的规则依次得出的：其中的一些相当于数学公理，而那种从一些公式推出另一些公式时所用的规则则相当于内容上的推理，从而内容上的推理就为那种由规则所决定的形式过程所代替。”[13] 这样，希尔伯特就把数学符号的发展推向了一个系统化、形式化的阶段。由于符号的形式化摆脱了对象具体的特定性，这就使数学研究的意义得到了极大的扩充。同时，符号的形式化又清楚地揭示了各种数学理论的逻辑结构，为深入研究各种数学理论的内在联系奠定了基础。

3. 符号方法的功能

(1) 明化数学问题，简化数学推理。

数学符号按其结构，可分为基本符号、组合符号和公

式符号。基本符号是表示单个概念的符号，它是符号系统中不可分割的最小单位，如 a、x、$+$ 等。若干基本符号的组合就形成组合符号。它表示复杂的数学概念，形成数学符号的"短语"，如 $(a+b)^2$ 等。如果组合符号与基本符号中的关系符号按一定规则相连，就形成公式符号，它表达一个判断一个命题。而多个公式符号的组成就形成数学中的推理。由于任何符号都是由基本符号组成的，而每个基本符号都是充分简缩的词或句子。故用符号表达的概念、判断、命题、推理自然要比自然语言简洁和明确得多，从而就有助于人们理解问题，分析和解决问题。正如怀特海所说的："这些术语和符号的引入，往往是为了理论的易于表述和解决问题。特别是在数学中，只要细加分析，即可发现符号化给数学理论的表述和论证带来极大的方便，甚至是必不可少的。"[4]

例如，在数学分析的发展史上，人们对证明变量的极限一直是束手无策的，变量能否到达它的极限也常争论不休而一无结果。然而，当极限概念采用 ε-δ 符号方式定义后，这些问题就迎刃而解了。正如世界著名数学家柯朗(R. Courant) 所说的："在 ε-δ 定义中独立变量并不运动；它在任何物理意义下并不'趋近于'或'迫近'极限 x_1。这些语句以及记号 $\rightarrow$ 依旧保留；它们所表达的富有启示性的直观感觉，是数学家不会也不应该放过的。但是在实际的科学过程中，欲验证极限的存在性，则必需应用 ε-δ 定义。这个定义是否圆满地反映了趋近概念的直观'动态'，这与几何公理是否提供对空间直观概念的圆满解释一样，它们属于同一类问题。两者的表述都舍弃了若干直观上实在的内容，但是在表示我们对这些概念的知识时，他们却成为适宜的数学结构。"[14] 可见，符号对揭示概念实

质及便于推理论证等都有极其重要的意义。

符号的形式是多样的，图形也是一种符号。希尔伯特说："算术符号是写出来的图形，而几何图形则是画出来的公式。"[4] 重视并运用"画出来的公式"，对明化数学问题、简化数学推理是十分重要的。法国著名数学家阿达玛(Hadamard)曾借用大数学家欧拉(Euler)的事例来提醒人们，他说："科学史上，欧拉所给出的下述例子是众所周知的。他为了向一个瑞典王子解释演绎的特性，而用圆来代表一般的概念，让我们考虑 A 和 B 两类事物，如果'凡是 A 都是 B'，则我们就想象圆 A 位于圆 B 之内；而如果'没有 A 是 B'，我们就想象圆 A 和圆 B 完全不相交；如果'某些 A 是 B'，则想象成圆 A 和圆 B 相交。我自己在进行某种推理时，也不用语言来代表思想，语言并不能使我看清这个推理过程是否正确，我使用类似于欧拉的方法，不过不是用圆圈，而是用某种不固定的图形来表示。"[15]

(2) 触发人们的创造性思维。

人的思维过程实际是一个对信息的处理、加工亦即组合，选择的过程，故进入人脑信息量的大小往往影响着人的思维质量。由于符号是高度浓缩信息的物质携带者，故应用符号思考常能缩减思维劳动，加速思维进程，从而易于获得创造性思维。尤其是符号的形式化发展，使得通过理性思维构思出某些新概念，常常成为数学发现的有力工具。拉普拉斯(Laplace)以数学分析的符号化对创造性思维的触发作用为例，论述道："数学分析的语言，是所有的数学语言中最完善的语言，而且语言本身就成为新发现的有力工具。特别是那些被构思出来的种种必要概念，往往是许多新算法的起点。"[4] 同时，由于符号常以直观、鲜明的形式将抽象的概念呈现在人们的眼前，故待

号思维往往具有简洁、明了、易为心灵感受的特点和优点，从而也有助于触发人们的创造性思维。

例如，由幂的运算公式 $x^m \cdot x^n = x^{m+n}$ 出发，可触发人们对分数指数幂的思考，因为 $x^{\frac{1}{2}} \cdot x^{\frac{1}{2}} = x$，故可用 $x^{\frac{1}{2}}$ 来记 $\sqrt{x}\,(x \geqslant 0)$；同样，由 $x^m \div x^n = x^{m-n}$，可创造出负指数幂和零指数幂的概念，牛顿正是出于这种触发才创造了分数指数幂和负指数幂。

(3) 实现思维的机械操作。

由于符号形式易于运算和推理，故研究问题时，人们可以暂时撇开符号的意义而仅着眼于形式，当符号与一定的概念单值地对应时，思想的操作可转换为对符号的操作，而符号的操作可委托机器进行，故人们利用符号借助计算机，便可使复杂、繁重的脑力劳动机械化，从而实现智力的解放。17 世纪时，莱布尼兹曾预言："一旦完成了这一计划，人类就将获得这样一种新工具，它对于人类理解能力的增强程度将远远超出任何一种光学仪器对于视力的加强。"[13]20 世纪 40 年代，电子计算机的出现使上述设想变为现实，它极大地解放了人们的智力劳动，其中"四色问题"的研究就是突出的一例。又因数学思维的机械化不但极大地拓展了数学的应用范围，而且还改变了数学应用的实践方式。比如天文学中超新星的爆发过程，地质学中的地壳运动，都难以在实验室里实验，却可以用计算机通过数学模型来模拟，从而对各种理论解释进行检验，因此，数学思维机械操作的出现，使科学研究的方式除了传统的理论工作和实验工作外，还出现了计算机实验。

(4) 促进数学和其他学科走向成熟。

一门学科的数学符号化程度，常常是这门学科是否成熟的重要标志。

数学史上不乏这样的事例，由于缺乏能够说清真正实质的符号，数学的某个领域就得不到发展。典型的例子是代数学，为了写出一般的代数方程式

$$a_0x^n + a_1x^{n-1} + \cdots + a_n = 0,$$

从丢番都到韦达和莱布尼兹用了整整13个世纪。在我国，宋代秦九韶解了许多高次方程(包括10次方程)，但没能发现一般二次方程的根与系数关系。宋元时期，由天元术发展到四元术而不能再向多元方向发展，致使四元术成为用算筹解方程组的顶峰。所有这些，都与中国古代缺乏先进符号有极大关系。正如卡约里(F. Cajori)在他的《数学符号史》中指出的："中国代数学在14世纪以后停滞不前的事实，主要是由于它不完善的，无适应性的符号。"可见，先进的符号对数学理论的建立和发展具有重要意义。

不仅数学科学的成熟发展依赖于先进的符号，就是其他学科的成熟发展也是如此。康德(Kant)说："我坚定地认为，任何一门自然科学，只有当它能应用数学工具进行研究时，才能算是一门发展渐趋完善的真实科学……而且一门科学对于数学工具的应用程度，就是这门科学渐变为真实科学的发展程度。"[4] 科学史上，正是由于牛顿、麦克斯韦、爱因斯坦、海森堡等人分别应用了微积分、偏微分方程、黎曼几何、矩阵等数学工具，才促进了物理科学从感性到理性的不断成熟。

4. 符号的选择原则

数学符号的发明史表明：一方面，数学符号随着社会实践的发展，不断淘劣存优，不断创新。其间，许多富有创

造精神的数学家，如韦达、笛卡儿、牛顿、莱布尼兹、希尔伯特等都作出了重要的贡献；另一方面，由于历史的和认识的局限性，在现存的符号体系中还存在若干令人遗憾的不足。对此，布尔巴基甚至说："只要粗略地检查一遍现存的符号，就表明很少有一看就懂十分简单的…… 其中大多数符号要不把它们复杂化到毫无用处的程度就不足以消除掉它的所固有的岐义。"[16]M. 克莱因也说："到17世纪末，数学里已特意(而不是偶然或碰巧地)使用符号并认识到它所能赋予的功效和一般性。只可惜那些并不认识符号重要性的人漫不经心而随意引入的符号太多了，而至今却都已通用 ……'我们今日所用的符号，是许多早已摒弃的符号系统中个别残留记号的杂烩'。"[17] 正是由于这样的原因，布尔巴基才把修订数学符号作为自己的一项重要工作。

数学符号作为数学科学中的语言符号，它的能指与所指间的关系，开始时是任意的，但一经选定就相对稳定。在诸多的能指与所指的关系中，怎样科学地选择具有约定俗成的关系，在组成组合符号、公式符号时，基本符号又该怎样选择，应遵循以下一些主要的原则。

(1) 整体性原则。由于数学的进一步社会化和形式化，使数学符号不断自我生成新的符号。这样，就有可能使符号的使用造成混乱、重复和失去控制。这就要求人们不断加强对符号的审视和修订，且在修订时要充分注意数学的系统性、逻辑性和各数学分支符号的一致性、协调性和无矛盾性，不要顾此失彼，前后相悖。例如，在数理逻辑中，否定、合取、析取、蕴涵、双向蕴涵，怀特海、罗素的记法是$\sim P$、$P\cdot Q$、$P\vee Q$、$P\supset Q$、$P\equiv Q$；希尔伯特的记法是$\overline{P}$、$P\&Q$、$P\vee Q$、$P\rightarrow Q$、$P\sim Q$；而波兰学者的记法

是N_p、K_{pq}、A_{pq}、C_{pq}、E_{pq}。几乎是一个作者一套符号，不但重复，而且相互矛盾。因而需要加以审视和修订，做到使用贯彻始终的统一符号，真正使符号系统成为统一、有序、相容的整体，以保证数学的整体性、逻辑性和系统性。这就是数学符号选择的整体性原则。

(2) 单义性原则。任何自然语言都有两个基本组成部分，一是元素(字母)和词汇，二是语法和语法结构。但自然语言无论是字母、词和语法，语法结构都是非单义的。其表现是一个字或词常有多种含义，且"自然语言中语法结构和逻辑结构之间不存在普遍性和必然性的一致……使得语言有可能歪曲和臆造思想"等。这种非单义的情况，严重妨碍了数学的逻辑推理和精确研究，故作为数学科学人工语言符号的数学符号必须是单义的。即符号的能指与所指的关系要互为单值映射，且用基本符号组合成各种组合符号和推理符号时，其语法必须按预定的法则进行。例如，在数学史上，角的符号与小于符号曾发生冲突，都用"$<$"表示。为满足单义性，前者后来改为"$\angle$"；数论中，高斯(Gauss)和勒让德(Legendre)分别用符号"$\equiv$"与"$=$"来表示同余，因符号"$=$"已用来表示一般的相等，故人们舍去了勒让德的符号而保留了高斯的符号。

(3) 简明性原则。要求符号及其系统都应为最简的。正如皮亚诺所指出的："数学的一切进步都是对引入符号的反应……在两种符号体系中，符号用得较少的一般是可取的。"例如，对于解决集合悖论，罗素和他的学生兰姆赛(Ramsay)都提出了解决的方案。罗素的方案是所谓的"分支类型论"。它不仅按对象的性质、性质的性质……加以分类，而且对每一类的性质，又按性质的意义

加以分级。在类中分级的基础上，利用恶性循环原则排除悖论。而兰姆赛的简单类型论是在废除分支类型论中关于级的划分后仅保留类的划分且承认类型混淆原则而不承认恶性循环原则的基础上建立起来的。由于简单类型论比分支类型论所涉及的符号系统要少得多，故简单类型论要比分支类型论简单得多，从而也就易于为人们所接受。此外，在基本符号的选择上，应使书写、排版方便，且“用词要尽可能地短，并且容易翻译成主要的科学语言”。在用基本符号组成公式符号时要使公式整齐、对称，运算简洁、明了。

例如，小数 2.5，至 19 世纪还有 $2\lfloor\underline{5}$；2'|5；2△5；2‘5；2,5 等记法，相比起来，目前用的记号书写要简便些。表达线段、角、三角形面积的加减，如选择有向线段、有向角、有向面积符号，就会使运算简洁，论证明了。射影几何中引进无穷远元素符号，就可以建立对偶原理。分式与根式采用幂的记号，就能化幂的高级运算为指数的低级运算。欧拉选用小写字母 a、b、c 表示三角形的边，其对应的大写字母 A、B、C 表示相应的角，就能使三角形诸公式整齐、对称，简单、明了，等等。所有这些都不失为是符号选择简明性原则的范例。

(4) 表意性原则。在可能的情况下，符号的选择应力求反应该符号所指概念、公式的思维特征。这主要包括概念、公式的由来、实质，客观现实原型，几何直观和隐含着的丰富、深刻的内容等。或如丢东涅(Dieudonne) 所说的，“如果可能的话”，要“能够让人想起它们所指称的概念或者它们的创始人”。例如导数符号 $\frac{dy}{dx}$ 就是能揭示导数、微分思维实质的一种符号。因为它表明，函数在某点

处的导数，等于函数在该点处的纵坐标改变量与横坐标改变量的商，亦即在该点处的切线斜率，这就深刻揭示了导数、微分概念的内在联系，并具有鲜明的直观形象。正如卡约里在评价这一符号时所指出的："尽管印刷不是特别合意，但这个符号现在比任何其他竞争者都得到更广泛的接受，最重要的原因是这个符号的适应性。应用简单的代数过程，人们就很容易从导数得到微分，给读者的心智直觉地暗示了带有两个垂直边 dy 和 dx 的直角三角形的特征。这个符号容易唤起达到微积分对几何和力学应用的核心思想"。

又如，用矩阵符号表达仿射变换

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = A\begin{pmatrix} x \\ y \end{pmatrix} + B。 \quad (*)$$

其中，$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$，$|A| \neq 0$，$B = \begin{pmatrix} a_{13} \\ a_{23} \end{pmatrix}$。

其思维特征揭示得更是淋漓尽致。事实上，设有仿射坐标系 $\{O; \vec{e}_1, \vec{e}_2\}$，它在仿射变换下的像是 $\{O'; \vec{e}_1{}', \vec{e}_2{}'\}$，则 $\vec{e}_1{}'$、$\vec{e}_2{}'$ 和 $\overrightarrow{OO'}$ 相对于 $\{O; \vec{e}_1, \vec{e}_2\}$ 的坐标分别是 A 的第一、第二列向量和 B，而 $|A|$ 又恰是图形在仿射变换下的变积系数。且公式(*)还隐含了"任何不共线的三对对应点决定唯一一个仿射变换"这一平面仿射几何基本定理。

(5) 科学性原则。符号指代的概念要正确、无误、严格、科学，没有潜在的歧义。符号指代什么，也必须明白、确切，不可模棱两可，似是而非。如长期以来，人们对 f 与 $f(x)$ 一直混淆不清，常常将依变元 y 亦即 $f(x)$ 误认为是函数关系。其实，f 是对应法则或算法，$f(x)$ 即 y 是 x 按对应法则 f 所得的函数值，$y = f(x)$ 是 y 与 x 间的函

数关系的一般表示。将 f 与 $f(x)$ 混为一谈，便违反了符号选择的科学性原则。

5. 符号方法的应用

(1) 符号逻辑。

符号逻辑是符号方法的深入和发展，也是符号方法应用的重要方面，它是以研究推理过程的形式结构和典范规律为其主要内容的，因而是分析、论断实际问题的有力工具。

设 P、Q 代表两个陈述或命题，符号逻辑中采用 —、$\wedge$、$\vee$、$\rightarrow$、$\leftrightarrow$ 分别表示“非”、“而且”、“或者”、“如果，……，则……”和“当且仅当”，则可将 p、Q 组成五个复合语句或命题。

$\overline{P}$，表示“非 P”，它是 P 的否定式；

$P \wedge Q$，表示“P 而且 Q”，是 P 与 Q 的合取式；

$P \vee Q$，表示“P 或者 Q”，是 P 与 Q 的析取式；

$P \rightarrow Q$，表示“如 P，则 Q”，称为蕴涵式，意谓有 P 就有 Q，亦即 P 是 Q 的充分条件，Q 是 P 的必要条件。

$P \leftrightarrow Q$，表示“P 当且仅当 Q”，称为双向蕴涵式。意谓 P 与 Q 互相蕴涵，亦即 P 是 Q 的必要且充分条件，反之 Q 也是 P 的必要且充分条件，双向蕴涵也可记成 $(P \rightarrow Q) \wedge (Q \rightarrow P)$。

若用 T(truth) 和 F(falsity) 分别代表真和假，则复合陈述或命题的真、假由如下真值表决定：

P	$\overline{P}$	P	Q	$P \wedge Q$	$P \vee Q$	$P \rightarrow Q$	$P \leftrightarrow Q$
T	F	T	T	T	T	T	T
F	T	T	F	F	T	F	F
		F	T	F	T	T	F
		F	F	F	F	T	T

以上真值表中其他各项是易于理解的。对 $P \to Q$ 的真值表的第一、三、四行，我们用以下的例子来说明它的直观背景，而 $P \to Q$ 真值表的第二行是十分显然的。

设有命题：

如果 $x > 5$，则 $x > 2$。 (＊)

在数学推理中，这是一个真的命题。其中 x 可取任意实数值。今分别取(1)$x = 6$；(2)$x = 3$；(3)$x = 1$，命题(＊)为：

(1) 如果 $6 > 5$，则 $6 > 2$；

(2) 如果 $3 > 5$，则 $3 > 2$；

(3) 如果 $1 > 5$，则 $1 > 2$。

(1) 是前件、后件均真，就是真值表中的第 1 行；(2) 是前件假、后件真，是真值表中的第 3 行；(3) 是前件、后件均假，是真值表中的第 4 行。

对上述逻辑运算符号，规定运算顺序依次为 —、$\wedge$、$\vee$、$\to$、$\leftrightarrow$(顺序的规定不唯一)，并可建立一些符号运算公式，如分配律：

$P \wedge (Q \vee R) = P \wedge Q \vee P \wedge R$，

$P \vee Q \wedge R = (P \vee Q) \wedge (P \vee R)$。

摩根律：

$\overline{P \wedge Q} = \overline{P} \vee \overline{Q}$，$\overline{P \vee Q} = \overline{P} \wedge \overline{Q}$。

符号逻辑可用于对实际和理论问题的分析与推断。

例 1 治安案例分析。

某市有一公寓，以前治安一直很好，但近期常发生失窃事件。警方经调查终于查明作案的罪犯是最近迁来的甲、乙、丙、丁四人，又查明(1) 甲从不单独作案；(2) 乙、丁都不单独作案，但他们分别同甲一起就要作案；(3) 丙是个惯偷，他一进公寓就要作案。问这公寓不发生失窃案

件时甲乙丙丁四人的情况。

设 T 表示公寓发生失窃事件。P、Q、R、S 分别表示甲、乙、丙、丁在公寓里，则按问题给定条件有命题公式

$$R \vee (P \wedge Q) \vee (P \wedge S) \to T,$$

应用合取对析取的分配律，得

$$R \vee P \wedge (Q \vee S) \to T,$$

则 $\overline{T} \to \overline{R \vee P \wedge (Q \vee S)}$，

用摩根律化简命题逻辑式，得

$$\overline{T} \to \overline{R} \wedge (\overline{P} \vee \overline{Q} \wedge \overline{S})。$$

故公寓里不发生失窃案件时，四人的情况是：丙不在这公寓，而且或甲不在公寓，或乙、丁都不在公寓。

例 2　$\lim\limits_{x \to a} f(x) \neq l$ 的 ε-δ 语言。

注意到在极限理论中，$\lim\limits_{x \to a} f(x) = l$ 的 ε-δ 语言是：

任给 $\varepsilon > 0$，一定存在一个 $\delta > 0$，只要 $|x - a| < \delta$，就有 $|f(x) - l| < \varepsilon$。亦可写成

$$\forall \varepsilon\{\varepsilon > 0 \to \exists \delta[\delta > 0 \wedge \forall x(|x - a| < \delta \to |f(x) - l| < \varepsilon)]\}$$

命题$\lim\limits_{x \to a} f(x) \neq l$ 是上述命题的否命题，若命题 P 的否记为 $\sim P$，则$\lim\limits_{x \to a} f(x) \neq l$ 即为

$$\sim \forall \varepsilon\{\varepsilon > 0 \to \exists \delta[\delta > 0 \wedge \forall x(|x - a| < \delta \to |f(x) - l| < \varepsilon)]\}$$

$$= \exists \varepsilon \sim \{\varepsilon > 0 \to \exists \delta[\delta > 0 \wedge \forall x(|x - a| < \delta \to |f(x) - l| < \varepsilon)]\}$$

$$= \exists \varepsilon\{\varepsilon > 0 \wedge \sim \exists \delta[\delta > 0 \wedge \forall x(|x - a| < \delta \to |f(x) - l| < \varepsilon)]\}$$

(因为 $\sim (P \to Q) = P \wedge \sim Q$)

$$= \exists \varepsilon\{\varepsilon > 0 \wedge \forall \delta \sim [\delta > 0 \wedge \forall x(|x - a| < \delta$$

$\to |f(x)-l|<\varepsilon)]\}$

$= \exists\, \varepsilon\{\varepsilon>0 \wedge \forall\, \delta[\delta \leqslant 0 \vee \sim \forall\, x(|x-a|<\delta \to |f(x)-l|<\varepsilon)]\}$

$= \exists\, \varepsilon\{\varepsilon>0 \wedge \forall\, \delta[\delta \leqslant 0 \vee \exists\, x \sim (|x-a|<\delta \to |f(x)-l|<\varepsilon)]\}$

$= \exists\, \varepsilon\{\varepsilon>0 \wedge \forall\, \delta[\delta \leqslant 0 \vee \exists\, x(|x-a|<\delta \wedge \sim (|f(x)-l|<\varepsilon)]\}$

$= \exists\, \varepsilon\{\varepsilon>0 \wedge \forall\, \delta[\delta \leqslant 0 \vee \exists\, x(|x-a|<\delta \wedge |f(x)-l| \geqslant \varepsilon)]\}$

$= \exists\, \varepsilon\{\varepsilon>0 \wedge \forall\, \delta[\delta > 0 \to \exists\, x(|x-a|<\delta \wedge |f(x)-l| \geqslant \varepsilon)]\}$

$(\because \sim P \vee Q = P \to Q)$

据最后一式知，$\lim\limits_{x\to a} f(x) \neq l$ 的含义是：

存在一个 $\varepsilon>0$，对一切 $\delta>0$，必定存在 x，虽然 $|x-a|<\delta$，但是 $|f(x-l)| \geqslant \varepsilon$。

(2) 符号方法的其他应用。

在数学问题的求解中，正确、合理使用符号可使求解过程得到简化。

例如，设有 9 只茶杯，杯口全部朝上，若将其中 6 只翻转过来，称为一次“运动”。问是否可经过有限次运动，使得茶杯的杯口全部朝下？

设将茶杯杯口朝上、朝下的状态分别用 $+1$、-1 来标记，这样，每次运动就是将其中 6 个数改变符号，亦即用 -1 分别去乘 6 个数。于是原问题就成为能否经过有限次运算将 9 个 $+1$ 全部变成 -1。

考虑 9 个数的乘积：由于每次运动相当于将 6 个数各乘以 -1，而 $(-1)^6=+1$，故不论经过多少次运动，9 个数的乘积将保持不变，从而不能把 9 个 $+1$ 变为 9 个

－1，也就不能使9只杯口全部朝上的茶杯变为杯口全部朝下。

上述思考中，我们用数字符号＋1、－1分别代表茶杯杯口向上和向下的状态，并用6个数的乘积来分析每次运动的结果，而获得了问题求解的创造性方法。这表明依据问题的具体特点和条件，引进特殊的符号进行分析，常可找到问题的创造性解法。

符号方法应用的关键在于对符号的含义，实质的确切理解。切忌含糊其词，一知半解。

例如，f 是对应法则，自变量 x 取值于正实数，且满足关系式

$$f(x)=f\left(\frac{1}{x}\right)\lg x+1,$$

求 $f(x)$。

在关系式 $f(x)=f\left(\frac{1}{x}\right)\lg x+1$ 中以 $\frac{1}{x}$ 替代 x，得

$$f\left(\frac{1}{x}\right)=f(x)\cdot\lg\frac{1}{x}+1,$$

所以

$$f(x)=f\left(\frac{1}{x}\right)\lg x+1=[-f(x)\lg x+1]\lg x+1$$
$$=-\lg^2 x f(x)+(\lg x+1),$$

从而

$$f(x)=\frac{1+\lg x}{1+\lg^2 x}。$$

本题求解的关键是对函数概念本质的理解，因为 $f(x)$ 表示自变量 x 在对应法则 f 作用下的函数值，故在已知关系式中以 $\frac{1}{x}$ 代 x，便求得了 $f(\frac{1}{x})$。又上例中我们还看到统一的幂次符号与对数符号的联合使用，使得在

求幂的对数时,可直接应用将幂指数拿到对数符号外的运算性质而便于运算。

历史上,函数(function)一词作为数学用语,最早出现在莱布尼兹 1692 年的论文中,现在使用的记号是欧拉首创的。他在 1734 年的著作中,曾用 $f(\frac{x}{a}+c)$ 来表示 $\frac{x}{a}+c$ 的函数。不过欧拉的函数概念没有明确指出对应法则。他说:"函数 $y=f(x)$ 是随手画出来的曲线,所表示的是 y 与 x 的关系。"对数符号,是对数发明者耐普尔(Napier)首先使用的,它是"logarithm(对数)"一词的缩写,首见于 1614 年耐普尔的著作《奇妙的对数定理说明》,耐普尔的著作发表后,引起英国数学家布里格斯(H. Briggs)的兴趣,他专程访问了耐普尔,并建议使用以 10 为底的对数,即为今天的常用对数。

以上我们看到了函数、对数、幂次符号的作用,它们均不失为优异符号的代表。与此相对照,中国在"辛亥革命"前不久,还用"⊥"、"T"分别表示加、减号,分子在分数线之下,分母在其上。字母用十字天干甲、乙、丙、丁、……、壬、癸和十二字地支子、丑、……、戌、亥表示英文中的 A、…、V;W、X、Y、Z 用天、地、人、元表示。如日本上野清著,中国徐宪臣译(1905 年付印)的《普通新代数学教科书》中,印有

$$\frac{\text{五}}{\text{丁}^{\text{二}}}\ \top\ \frac{\text{三}}{\text{丙}^{\text{二}}}\ \perp\ \frac{\text{二七}}{\text{甲}^{\text{二}}\text{乙}^{\text{二}}}$$

即 $\frac{d^2}{5}-\frac{c^2}{3}+\frac{a^2b^2}{27}$[18]。

这样的符号怎样能适应迅速发展的数学和科学。

例 3 换位问题。

设有 49 位同学,坐成 7 行 7 列,若每一座位的前、后、

左、右座位叫做它的"邻座"，问要让这49位同学中的每一位都换到他的邻座上去，这样的换位方案能否实现？

问题中，每一座位有前、后、左、右四个邻座，邻座又有邻座，情况复杂，一时难于入手。但若借用符号方法思考，脉络就清楚了。设用 A 与 B 标志这49个座位：

$$
\begin{matrix}
A & B & A & B & A & B & A \\
B & A & B & A & B & A & B \\
A & B & A & B & A & B & A \\
B & A & B & A & B & A & B \\
A & B & A & B & A & B & A \\
B & A & B & A & B & A & B \\
A & B & A & B & A & B & A
\end{matrix}
$$

此时，对于标 A 的座位只有标 B 的才可能是它的邻座，反过来亦如此。故换位就是要使 A 的座位换到某些 B 上去，B 的座位换到某些 A 上去，这实际就是 A 与 B 的座位之间的一一对应问题。由于两种座位的总和49是奇数，则每种座位的数目必一奇一偶地数目不等，因而它们间不能建立一一对应关系，故问题中调位方案不能实现。

例4　若干纯数学问题的符号使用及其要点。

① 符号的使用要以指代的对象存在为前提。

证明：如果

$a_1 = \sqrt{2}$,

$a_2 = \sqrt{2 + \sqrt{2}}$,

……

$a_n = \underbrace{\sqrt{2 + \sqrt{2 + \sqrt{2 + \cdots}}}}_{n个}$,

……

则有$\lim\limits_{n\to\infty}a_n = 2$。

证明：设数列$\{a_n\}$的极限为a，则

$$a_{n+1} = \sqrt{2+\sqrt{a_n}}。$$

取极限

$$\lim_{n\to\infty}a_{n+1} = \lim_{n\to\infty}\sqrt{2+a_n},$$

有

$$a = \sqrt{2+a}。$$

解此方程，有

$$a = 2, a = -1(舍去),$$

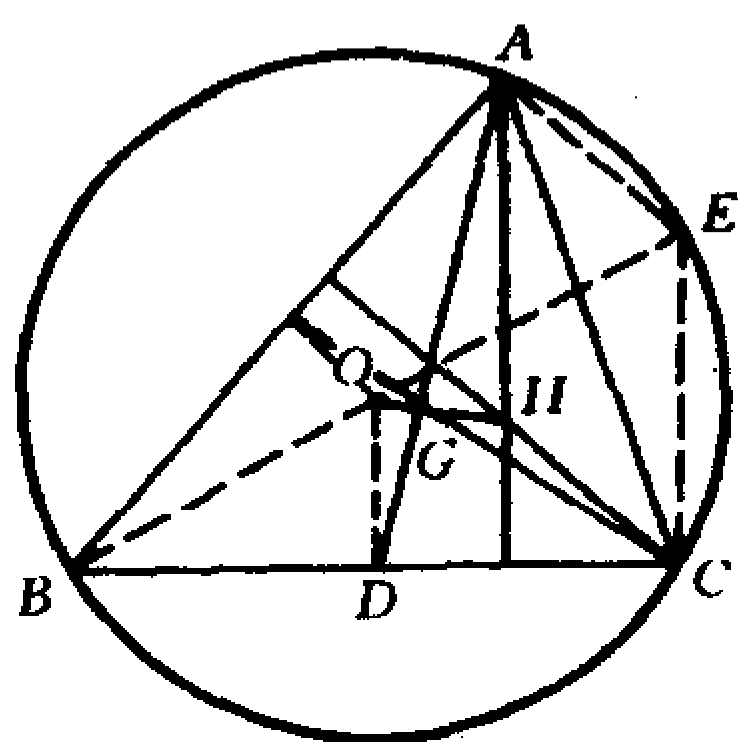
故

$$\lim_{n\to\infty}a_n = 2。$$

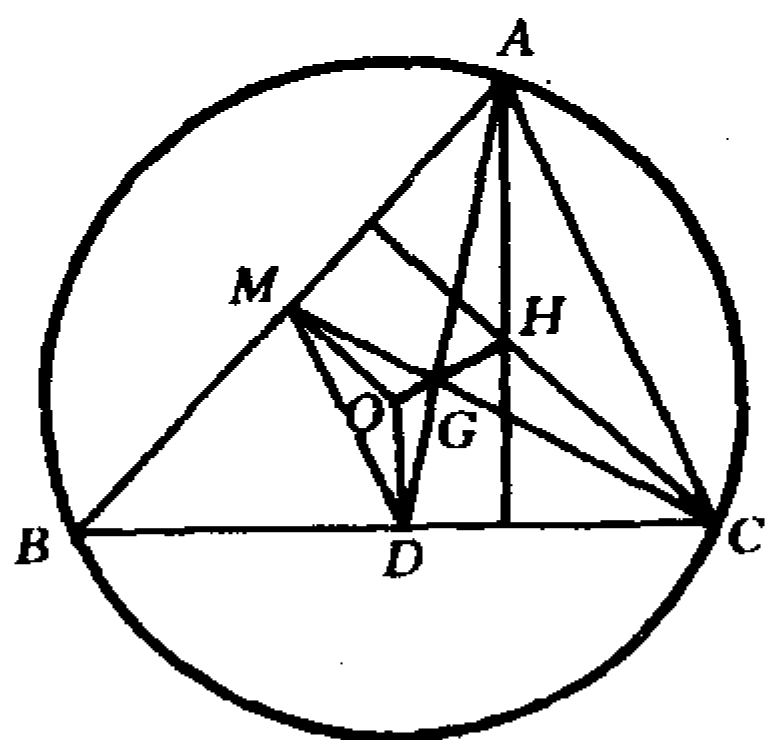
这里，符号a指代的对象的存在是题中首先要证明的，解题中，引进了尚需证明其存在的指代对象的符号，在逻辑上出现了循环论证。

② 符号使用的特殊和一般。

问题1：证明$\triangle ABC$的外心O、重心G、垂心H为一直线(欧拉线)。



(1)



(2)

图 2.1

证明：1°选用向量符号。设 BO 交 $\triangle ABC$ 外接圆于点 E，则 $AHCE$ 为平行四边形(两组对边对应平行)(图 2.1(1))。

又 $\quad \overrightarrow{OB}+\overrightarrow{OC}=2\overrightarrow{OD}=\overrightarrow{EC}$,

所以 $\quad \overrightarrow{OB}+\overrightarrow{OC}=\overrightarrow{AH}$，故

$$\overrightarrow{OH}=\overrightarrow{OA}+\overrightarrow{AH}=\overrightarrow{OA}+\overrightarrow{OB}+\overrightarrow{OC}。$$

但 $\quad \overrightarrow{OG}=\frac{1}{3}(\overrightarrow{OA}+\overrightarrow{OB}+\overrightarrow{OC})$，于是，$\overrightarrow{OH}=3\overrightarrow{OG}$,

即 O、G、H 共线，且 $OG:GH=1:2$。

2°选用无穷远元素符号，规定平面内一组平行直线确定唯一的无穷远点，所有无穷远点的集合形成一条无穷远直线(图 2.1(2))。

由于 $OM \parallel HC, DO \parallel AH, MD \parallel CA$，亦即

$OM \times HC = P_\infty, DO \times AH = Q_\infty$,

$MD \times CA = R_\infty$。

在三点形 DOM 和三点形 AHC 中，因 $P_\infty Q_\infty R_\infty$ 共线，由笛沙格逆定理知 AD、HO、CM 共点于 G，亦即 O、G、H 共线。

问题 2：已知三角形一边的平方等于第二边乘以第二、第三边的和，则一边的对角等于第二边对角的 2 倍。

证明：引进符号。用 $\triangle ABC$ 表示已知三角形，a、b、c 表示角 A、B、C 的对边，则原问题化为：

已知 $a^2=b(b+c)$，求证：$\angle A=2\angle B$。

由 $A+B+C=180°$，知 $\sin C=\sin(A+B)$，故

$a^2=b(b+c) \Rightarrow \sin^2 A=\sin B(\sin B+\sin C)$

$\Rightarrow \sin^2 A=\sin B[\sin B+\sin(A+B)]$

$\Rightarrow \sin^2 A-\sin^2 B=\sin B\cdot\sin(A+B)$

$\Rightarrow \sin(A+B)\cdot\sin(A-B)=\sin B\cdot\sin(A+B)$。

$\because \quad \sin(A+B) \neq 0, \therefore \quad \sin(A-B) = \sin B$。

又 $A - B \neq 180° - B$，故 $A - B = B$，从而 $A = 2B$。

符号的使用，既有一般性又有特殊性。如涉及三角形边、角关系的定量问题，通常用大写字母表示角，相应的小写字母表示边。这样表示，使三角形中的诸公式、定理都获得了对称、明了的形式。如正弦定理

$$\frac{a}{\sin A} = \frac{b}{\sin B} = \frac{c}{\sin C},$$

余弦定理

$$c^2 = a^2 + b^2 - 2ab\cos C,$$

面积公式

$$S = \frac{1}{2}ab\sin C = \frac{1}{2}bc\sin A = \frac{1}{2}ac\sin B$$

等，问题 2 的符号使用正是采用了这种一般方法。

符号使用还需注意特殊性，在不同问题、不同思路的情况下，使用的符号往往相异。如问题 1 分别采用了向量符号和无穷远元素符号。相比而言，后者证法简洁，但前者却给出了三心位置的定量刻画，是后者所不及的。

2.2 对偶思维方式

对偶思维方式是数学中实现思维经济的重要手段，它对数学理论的探索和数学原理的刻画具有重要的意义。

1. 对偶性

两个概念是“对偶的”，是指两个具有相互依存、相互制约关系的对象或现象的概念，它们通常是一一对应、成对存在的。对偶性是事物的基本属性，因为任何事物往往都处于方向相反、作用不同的对立状态之中，且双方相互

联结，相互依赖，各方都以对方为自己存在的条件。应用对偶着的双方相互依存、成对存在的关系来分析思考问题的方式，就是对偶的思维方式。

数学中，正负数、共轭复数、亲和数(彼此等于对方所有因数之和的一对正整数)、孪生素数(差为 2 的相邻两素数对)、互逆运算、互逆变换等都是由事物的对偶性所引出的研究课题。其中有些研究课题经久不衰，如孪生素数对问题，人们已获知在小于 10^5 自然数中有 1224 对，在小于 10^6 中有 8164 对，在小于 3.3×10^7 中有 152892 对。有人猜测孪生素数对的个数是无限的，但至今却不能证明。可见对偶性在数学思维中的重要性。

2. 对偶原理

在数学中，某些对偶的事物，虽然本身意义不同，但其抽象的规律或性质，不仅可以一一对应，而且也可能完全一致。这样，就有可能使具有这种性质的两对偶对象，建立起如下的结构关系体系，使在该体系中，对某一对象成立的命题，对其对偶的对象同样也成立，这时就说该体系实现了结构关系的对偶化，并说它们间建立了对偶原理。

(1) 射影几何中的对偶原理。

射影平面上的点和直线，若将点视为基本元素时，直线就是点的轨迹；反过来，当直线作为基本元素时，点就是直线的包络。因此，射影平面上的点和直线，就其意义和作用而言，是方向相反的一对，它们相互联结，相互制约。射影平面上任一点的齐次坐标(x_1,x_2,x_3)(x_i 不全为零)，与直线的齐次坐标(u_1,u_2,u_3)(u_i 不全为零) 不仅形式相似，而且“直线通过点”与“点在直线上”都表现出同一种抽象规律

$$(x_1, x_2, x_3) \cdot (u_1, u_2, u_3) = 0,$$

即

$$u_1x_1 + u_2x_2 + u_3x_3 = 0。$$

这样，凡关于点和直线结合关系的任一命题，当将点与直线对换时得到一个新命题。但这种对换，丝毫也没有改变抽象规律的本质，因而新命题与原命题具有相同的真假性。故射影平面上的点与直线就结合性、顺序性等关系可以实现对偶化，即射影几何中对偶原则成立。

例如，笛沙格(G. Desargues) 正逆定理[1] 与[1]* 是对偶的，且同为真：

[1] 如果两个三点形对应顶点的连线共点，则它们对应边的交点共线；

[1]* 如果两个三线形对应边的交点共线，则它们对应顶点的连线共点。

这里，“三点形”与“三线形”对偶，“对应顶点的连线”与“对应边的交点”对偶，“共点”与“共线”对偶。知道了这些关系，不难从[1]得[1]*，或从[1]* 得[1]。需要说明的是，“三点形”与三线形”前者着眼于点，点作为基本元素，后者着眼于线，线作为基本元素。其实是一回事，均指三角形，不过这里所说的三角形，不是欧氏平面的三角形，而是射影平面的三角形，其顶点允许为无穷远点。

又如，非退化二次曲线既可作为点的轨迹的点曲线，又可作为直线包络的线曲线，与一个简单六点形和非退化二次曲线内接的问题相对应的，就是一个简单六线形与非退化二次曲线外切的问题。如下的[2] 和[2]* 也是对偶的，且同为真：

[2] 巴斯加(B. Pascal) 定理：内接于非退化二次曲线的简单六点形三对对边的交点共线。

[2]* 布列安桑(C. J. Brianchon) 定理：外切于非退化二次曲线的简单六线形三对对顶的连线共点。

这里的简单六点(线) 形是指六个(条) 点(线)，其中无三者共线(点)，及它们顺次两两的连线(交点) 所成的图形。

(2) 布尔代数中的对偶原理。

如果一个集合 M 中至少含有两个不同的元素 e_1, e_2，并且在其上定义了一个一元运算“$'$”和两个二元运算 $\oplus, \odot$，这三种运算在 M 上都封闭，且满足如下的运算规律：

1° $(\forall x)(\forall y)\ [x \oplus y = y \oplus x]$,

$(\forall x)(\forall y)\ [x \odot y = y \odot x]$;

(两个运算是可交换的)

2° $(\forall x)(\forall y)(\forall z)\ [x \odot (y \oplus z) = (x \odot y) \oplus (x \odot z)]$,

$(\forall x)(\forall y)(\forall z)\ [x \oplus (y \odot z) = (x \oplus y) \odot (x \oplus z)]$;

(运算 $\oplus$、$\odot$ 的任何一个，关于另一个是分配的)

3° $(\forall x)\ [x \oplus e_1 = x]$, $(\forall x)[x \odot e_2 = x]$;

(e_1 是关于 $\oplus$ 的单位元，e_2 是关于 $\odot$ 的单位元)

4° $(\forall x)(\exists x')\ [x \oplus x' = e_2]$,

$(\forall x)(\exists x')\ [x \odot x' = e_1]$。

(符号 $\forall$、$\exists$ 分别表示“任意”和“存在”)

则称 M 为一般的布尔代数系统。在布尔代数中，如果将定理中的 $\oplus$、$\odot$、e_1、e_2 依次更换为 $\odot$，$\oplus$，e_2，e_1，得到的定理是原定理的对偶定理，若原定理为真，则对偶定理亦真，即对偶原理在布尔代数中成立。

例如：$A \oplus (A' \odot B) = A \oplus B$ 的证明是：

$$A \oplus (A' \odot B) \overset{2^\circ}{=} (A \oplus A') \odot (A \oplus B)$$
$$\overset{4^\circ}{=} e_2 \odot (A \oplus B)$$
$$\overset{1^\circ}{=} (A \oplus B) \odot e_2$$
$$\overset{3^\circ}{=} A \oplus B。$$

其对偶定理为

$$A \odot (A' \oplus B) = A \odot B,$$

其证明过程也只要将 $\oplus$、$\odot$、e_1、e_2 相应换成 $\odot$、$\oplus$、e_2、e_1 即可。

$$A \odot (A' \oplus B) \overset{2^\circ}{=} (A \odot A') \oplus (A \odot B)$$
$$\overset{4^\circ}{=} e_1 \oplus (A \odot B)$$
$$\overset{1^\circ}{=} (A \odot B) \oplus e_1$$
$$\overset{3^\circ}{=} A \odot B。$$

(3) 线性规划中的对偶原理。

线性规划中的一个最大(小)值问题,相应地存在着一个特定的包含同样数据的最大(小)值,用数学语言刻画,即两个线性规划 LP 与 LP^*(LP^* 即 Linear programming,“线性规划”简称):

LP	LP^*
$\max S = C^T X$	$\min W = B^T Y$
$AX \leqslant B$	$A^T Y \geqslant C$
$X \geqslant 0$	$Y \geqslant 0$

其中,C、X 是 $n \times 1$ 矩阵,Y、B 是 $m \times 1$ 矩阵,A 是 $m \times n$ 矩阵,T 为矩阵转置符号。

若 LP 与 LP^* 之中一个有解,且这个解是有限的,则另一解也存在,它也是有限的,而且 $\max S = \min W$。

对偶原理除以上所述内容外,在偏序集、布尔代数、格论、范畴论、对策论和伽利略非欧几何中也都成立,这

里不再赘述。

3. 对偶思维的渊源及其功能

历史上，对对偶思维的重视起始于射影几何。19世纪，几何学家们发现，涉及平面图形的定理，如果把"点"换成"线"，"线"换成"点"，重述一遍，不但话说得通，而且同样正确。法国数学家彭赛列(J. V. Poncelet)首先对此作了研究，他将这种现象的原因归结于二次曲线的极点和极线间的一种对应关系。在1882年的《论图形的射影性质》和在1824年的《论配极的一般理论》中，他给出了极点和极线相互变换的一般表述，并以此建立了许多定理。不过，他这样建立的对偶原理需要一个圆锥曲线来作"中介"，具有某种局限性。突破这种局限性的是吉尔岗尼(J. D. Gergonne)。他认为，对偶原理是个普遍原理，适用于除涉及度量性质外的一切陈述和定理，不需要什么中间支撑物。"对偶性"一词就是首先由他引进的，用以表示一定理和由此对换点、直线得出的新定理间的关系。他还发明了将成对偶的定理分写在左右两旁的方法，一直沿用至今。但吉尔岗尼没有给出他的观点的充分理由和有力证据。后来，几何学家们展开的讨论完全弄清了这一问题，尤其是普鲁士的几何学者普吕克(J. Plücker)于1829年创立了线坐标，从而为对偶原理给出了精妙的代数证明。这样，对偶不依赖于中介的圆锥曲线的问题也就相应获得解决。在对偶原理认识的历史上，有意识地通过使用对偶原理来发展射影几何的当推瑞士数学家斯太纳(J. Steiner)，他把圆锥曲线对偶化，得到了一种称为线曲线的新结构而发现了大批定理，从而使他有幸成为历史上有过伟大贡献的综合几何学者。值得指出的是，19世纪末至20世纪初发展起来的公理化方法，为人们从自对

偶公理组的角度来建立射影几何体系从而获得这一原理的逻辑证明，对加深对偶原理的认识其作用是不可估量的。

对偶原理对射影几何发展的重要推动作用促使人们思考：在其他数学领域能否建立相应的原理？后来，在格论、布尔代数、偏序集、范畴论、线性规划、伽氏非欧几何等数学分支中都建立起了对偶原理。数学发展的历史表明，借助对偶方法，我们不仅可以得到新概念、新对象，而且还可以得到新定理（对象的属性和联系等）。因此，对偶方法有发展数学理论的功能，从而它具有真善的特点。由于对偶明显地具有简化理论的功能，正如数学家 Jrri Adamek 以范畴为例所指出的“对偶范畴的概念的重要性在于，它使得我们有可能实现关于范畴的每一个概念和定理的‘双化’”。[19] 因而它又具有美的特点，故对偶方法是一个真善美的统一体。

4. 对偶方法的应用

(1) 应用事物的对偶性进行数量关系的分析。

在处理数量关系的问题时，应用事物的对偶性分析问题，往往可使问题得到简化。

例如，一环形公路有若干汽车站，车站的海拔高度仅有 15 米、20 米两种。若这环形公路上水平公路与有坡公路的段数相等，则车站的数目必是 4 的倍数。

因为公路是环形的，所以车站数必等于公路的段数。又公路的水平段数等于有坡的段数，设其段数为 k。则环形公路的总段数为 $2k$，注意到有坡公路由上坡段、下坡段组成，故有坡公路段数必为偶数，即 $k = 2n$。所以，总段数

$$2k = 2(2n) = 4n,$$

它是 4 的倍数，从而车站数目亦是 4 的倍数，故问题得证。

平时在日常生活中，利用对偶概念思考问题的例子是很多的。例如，“上楼”和“下楼”是对偶的。一人上楼时跨过某一台阶，那么他下楼时必须还得跨过这一台阶。因此，这人上楼时跨过的台阶数必等于下楼时跨过的台阶数。从而他跨过楼梯的阶数必为偶数。这个事实的抽象就是下面的计数定理：设 A、B 两个有限集合，如果其元素间能建立起一一对应，则这两集合元素的和必为偶数。

以上就是利用对偶思想通过有坡公路段的上坡段和下坡段的对偶性推得其段数为偶数，从而证得了结论。

类似于本题的方法，还可证明：若自然数 n 不是平方数，则 $\sqrt{n}$ 的全体正因数的个数必是偶数。这是因为，n 不是平方数，则 $\sqrt{n}$ 必不是正整数，又设 k 是 n 的一个正因数($n=kl$)，则或者 $k>\sqrt{n}$，或者 $k<\sqrt{n}$。

当 $k>\sqrt{n}$ 时，$l=n/k<n/\sqrt{n}=\sqrt{n}$；

当 $k<\sqrt{n}$ 时，$l=n/k>n/\sqrt{n}=\sqrt{n}$。

这说明，如果 n 有一个大于 $\sqrt{n}$ 的正因数，那它必然也有一个小于 $\sqrt{n}$ 的正因数，即 n 的大于 $\sqrt{n}$ 与小于 $\sqrt{n}$ 的正因数可配对，于是知 n 的正因数个数是偶数。

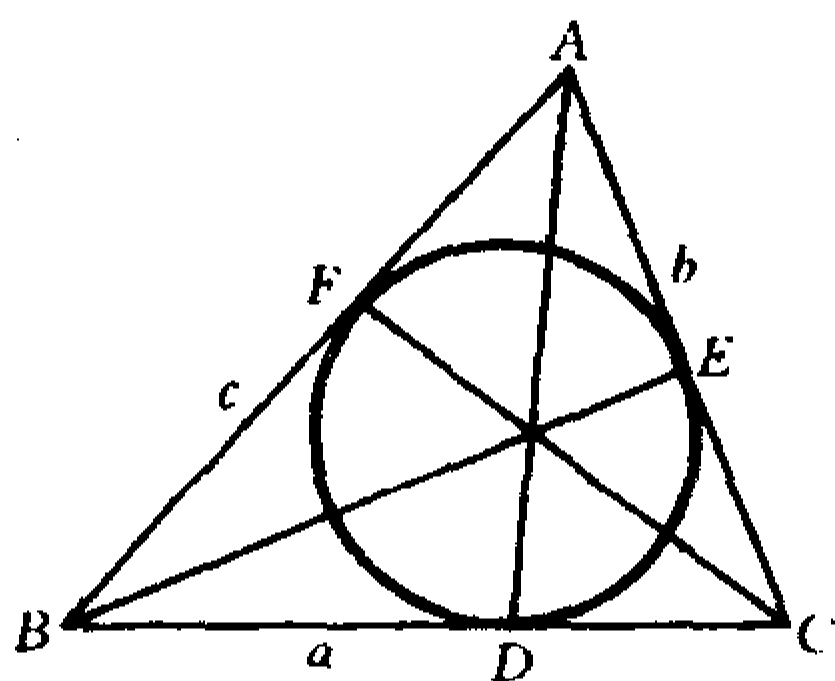
(2) 利用对偶原理探索未知定理和求解问题。

例 1 探索几何新定理。

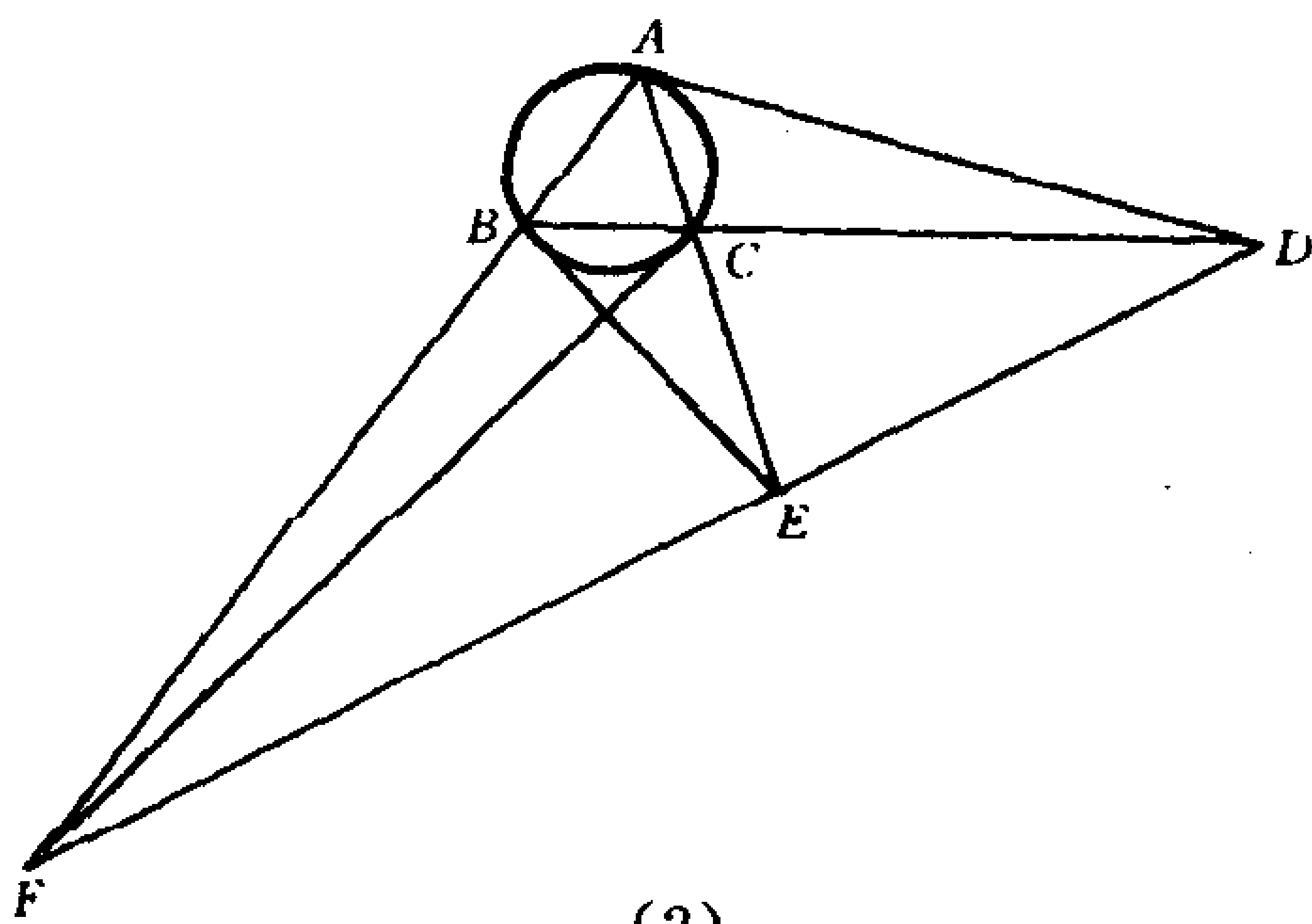
射影几何中的对偶原理可以帮助我们探索未知的几何定理。比如，已知定理：“三角形各顶点与该三角形内切圆在对边切点的连线共点”〔1〕(图 2.2(1))。

这时，我们应用对偶原理，将“圆内切于三角形”的关系对偶为“圆外接于三角形”，“顶点”换为“边”，“对

边”换为“对顶”,“切点”换为“切线”,“共点”换为“共线”,便得对偶命题:“三角形各边与该三角形外接圆过对顶的切线的交点共线。”〔1〕*(图 2.2(2))



(1)



(2)

图 2.2

已知前一命题为真,故本命题亦为真,从而获得了新定理。在后一定理的结论中,共线的三点中,可能都是有穷远点,也有可能有一个或两个(从而三个)无穷远点。显然,内接三角形在等腰时,三点中有一为无穷远点;在等边时,三点均为无穷远点。

事实上,〔1〕是布列安桑定理的特例。这时圆的外切六线形 *aabbcc* 退化成外切三线形,从而对顶连线便退化

为顶点与对边切点的连线，应用布列安桑定理便得

$bc \times aa = AD$、$ca \times bb = BE$、$ab \times cc = CF$ 共点。

而〔1〕* 则是巴斯加定理的特例。因为这时的圆内接六点形为 $AABBCC$，则 AA、BB、CC 便是过 A、B、C 的圆的切线。这样，$AA \times BC = D$、$AC \times BB = E$、$AB \times CC = F$ 三点共线，其中，AA、BB、CC 是过 A、B、C 的圆的切线。

例 2　空间线、面位置关系的内部联系。

立体几何中直线与直线、直线与平面、平面与平面的位置关系，内容众多，初学者不易掌握。如何从内容众多的几何定理中理出头绪，能否使用类似于对偶的方法，注意到一直线的方向可用其方向矢量表征，而平面的方向也可借助其法矢量表征。设以非零矢量$\vec{a}$、$\vec{b}$ 表示空间直线的方向矢量，或平面法矢量，则矢量等式

$$\vec{a} = k\vec{b}(k \neq 0), \qquad ①$$

$$\vec{a} \cdot \vec{b} = 0, \qquad ②$$

几何意义是：当$\vec{a}$、$\vec{b}$ 均是直线的方向矢量，或平面的法矢量时，① 表示两直线平行（含重合）或两平面平行（含重合）的充要条件；② 表示两直线垂直或两平面垂直的充要条件。而当$\vec{a}$、$\vec{b}$ 一矢量是直线的方向矢量，另一是平面法矢量时，① 是直线与平面垂直的充要条件，② 是直线与平面平行（含直线在平面内）的充要条件。亦即矢量等式 ① 和 ② 包含了直线与平面相互关系的各种情况。据此，可设计对偶原理，注意到对于某一矢量等式，当改变其矢量含义，例如$\vec{a}$ 原代表直线方向矢量现改为平面法矢量时，则在几何上相当于将此直线改为平面，且同时将与此直线的平行垂直关系改为与调换后的平面的垂直平行关系。从而有：

若将两重合直线(平面)看成是两平行直线(平面)的特殊情况,把直线在平面内看成直线与平面平行的特殊情况,则直线、平面的垂直、平行关系中存在着如下对偶原理:

把命题中某一直线(平面)换以平面(直线),同时把与这一直线(平面)有关的平行(垂直)关系换以垂直(平行)关系,所得的命题同真伪。

例如:设 l_1、l_2、l 为三直线,α 为一平面。

$\left.\begin{matrix} l_1 \parallel l \\ l_2 \parallel l \end{matrix}\right\} \Rightarrow l_1 \parallel l_2$ —— 直线 l 换成平面 α / 与 l 相关的"平行"换成"垂直" ——→ $\left.\begin{matrix} l_1 \perp \alpha \\ l_2 \perp \alpha \end{matrix}\right\} \Rightarrow l_1 \parallel l_2$

两命题同真。

又如:

点 M 为平面 β 外一点,则有且仅有一平面 α 使 $\alpha \parallel \beta$	平面 β 换成直线 l ——→ 与 β 相关的"平行"换成"垂直"	点 M 为直线 l 外一点,则有且仅有一平面 α 使 $\alpha \perp l$

两命题亦同真。

例 3 生产中的最大利润。

设某工厂安排生产 Ⅰ、Ⅱ 两种产品,这些产品分别要在甲、乙、丙、丁四种不同的设备上加工,按工艺要求产品 Ⅰ、Ⅱ 在各设备上需要加工的台时数及各台设备在一个计划时期内提供的有效台时数,如下表:

产品＼设备	甲	乙	丙	丁
Ⅰ	2	1	4	0
Ⅱ	2	2	0	4
有效台时	12	8	16	12

该工厂每生产一件产品 Ⅰ 可得利润 2 元,每生产一件产品 Ⅱ 可得利润 3 元。问应如何安排生产计划,才能

获得最多利润?若工厂考虑不进行生产而把全部可利用资源出租给其他企业,怎样确定合理价格?

解:用 x_1、x_2 分别表示产品 Ⅰ、Ⅱ 的产量,S 元为利润,则线性规划模型为:

$$LP: \max S = C^T X$$
$$AX \leqslant B$$
$$X \geqslant 0$$

其中,$C = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$,$X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$,$A = \begin{bmatrix} 2 & 2 \\ 1 & 2 \\ 4 & 0 \\ 0 & 4 \end{bmatrix}$,$B = \begin{bmatrix} 12 \\ 8 \\ 16 \\ 12 \end{bmatrix}$。

亦即:$\max S = 2x_1 + 3x_2$,

$$\begin{cases} 2x_1 + 2x_2 \leqslant 12, \\ x_1 + 2x_2 \leqslant 8, \\ 4x_1 \leqslant 16, \\ 4x_2 \leqslant 12, \\ x_i \geqslant 0 \ (i = 1,2)。 \end{cases}$$

现考虑从相反角度提出问题。设 y_1、y_2、y_3、y_4 分别表示设备甲、乙、丙、丁每台时的租赁价格。由于每生产一件产品 Ⅰ 可得利润 2 元,故将产品 Ⅰ 的设备台时用于外加工时所得的加工费不应少于 2 元,否则宁可生产 Ⅰ,于是有

$$2y_1 + y_2 + 4y_3 \geqslant 2。$$

同样有

$$2y_1 + 2y_2 + 4y_4 \geqslant 3。$$

用于外加工时总收入

$$W = 12y_1 + 8y_2 + 16y_3 + 12y_4,$$

且定价时不能任意抬高价格,故在满足大于或等于生产

Ⅰ、Ⅱ 产品利润的条件下，应尽量降低价格，才会有单位来加工，所以要求 W 有最小值，于是得到

$$\min W = 12y_1 + 8y_2 + 16y_3 + 12y_4,$$

$$\begin{cases} 2y_1 + y_2 + 4y_3 \geqslant 2, \\ 2y_1 + 2y_2 + 4y_4 \geqslant 3, \\ y_j \geqslant 0 \ (j = 1,2,3,4)。 \end{cases}$$

即 LP^*：$\min W = B^T Y$

$$A^T Y \geqslant C$$

$$Y \geqslant 0$$

其中，$Y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix}$，$A$、$B$、$C$ 同上。

显然 LP 与 LP^* 为对偶线性规划。

LP 问题在点(4,2) 上有最优值，其最优值为 $\max S = 2 \times 4 + 3 \times 2 = 14$(图 2.3)，故 LP^* 的最优值 $\min W = 14$。

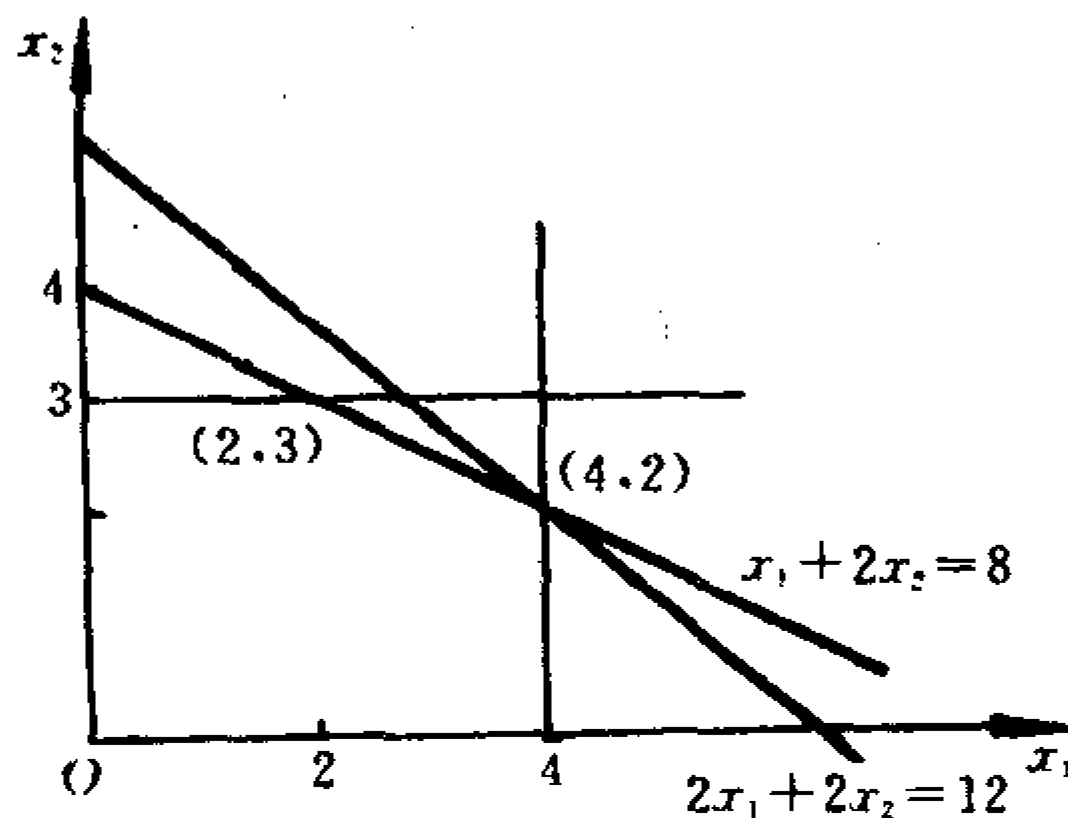


图 2.3

2.3　构造性思维方式

构造性思维方式是数学中一种重要的创造性思维方式，它对数学理论的创造、发展和数学问题的解决都具有重要意义，因为构造本身就是一种从无到有的创造。

1. 构造性方法的含义

通常数学中的所谓构造性方法，是指概念和方法按固定的方式在有限步骤内能进行定义或得以实现的方法。这种方法的基本特征是：

1°描述的直观性。即对讨论的对象能进行较为直观的描述。

2°实现的具体性。即不只是判明某种解的存在性，而且要实现具体求解。

与非构造性方法相比较，如同美籍华人王浩教授所指出的，前者是“做的数学”(mathematics of doing)，后者是“在的数学”(mathematics of being)[5]。

例如，中国古算中，用“更相减损术”求“等”的方法就是一种构造性方法。这里的“等”，即是两整数的最大公约数。其求法，《九章算术》“约分术”指出：“以少减多，更相减损，求其等也。”如 24 与 15 的等是 3，就是按下述方法求得的：

$$[24,15]\to[9,15]\to[9,6]\to[3,6]\to[3,3]$$

其原理是，在运算过程中，整数逐步减小，但“等”却始终不变。由于这种“更相减损”术，总可在有限步“更相减损”的过程中把等求出，故它是一种构造性方法。

又如，求一元二次方程 $ax^2+bx+c=0$ 的根，可按求根公式在有限步骤内通过四则运算和乘方、开方运算

将它求出来，故它亦是构造性方法。

与这种方法相区别的，如连续函数性质的中间值定理："设 $f(x)$ 在 $[a,b]$ 上连续，并且 $f(a)\cdot f(b)<0$，则存在一个 $c(a<c<b)$ 使得 $f(c)=0$。"数学分析中的证明方法是：通过反证法证明 c 的存在性，它并没有给出一个能行过程使之在有限步骤内将这个 c 的存在确定地构造出来，故这样的方法是一种非构造性方法。

2. 构造性方法的渊源与作用

构造性方法自数学产生之日起就相伴而产生了。但将它明确提出，以至把它推向极端，并致力于这一方法研究的，与数学基础的直觉派有关。直觉派出于对数学"可信性"的考虑，提出了诸如"存在必须是被构造"等口号，主张"定义应当包括由有限步骤所定义的对象的计算方法，而存在性的证明对于要确立其存在的那个量，应当许可计算到任意的精确度"。[17] 亦即认为"定义和证明都必须是构造性的"。诚然直觉派在哲学观点上是片面的，但他们对数学中构造性方法的具体研究当加以肯定。近代数学构造性方法经历从直觉数学到算法数学再到现代构造数学的发展，离不开直觉派的努力。

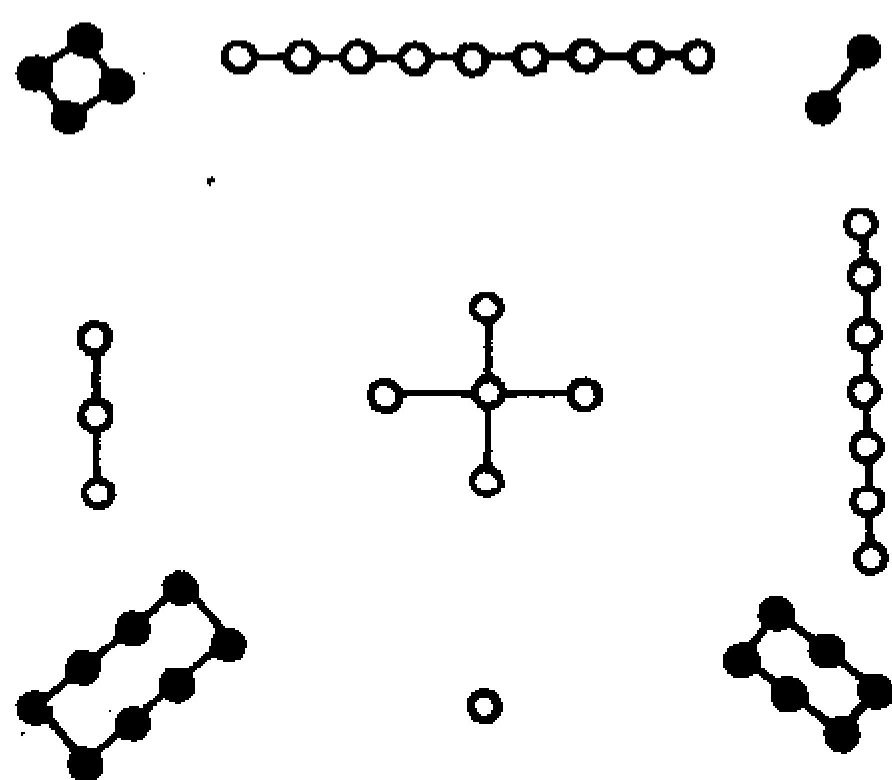
构造性方法的作用，突出地表现在它的创造价值和应用价值上。因为它"侧重于思维的构造性实践"，亦即注重于获得结果并将取得的结果构造出来。因而具有明显的应用价值，尤其现代计算机的发展，使这种应用价值更为突出。同时，要获得种种结果的构造绝非易事，它本身就是一种创造。故构造性方法具有突出的创造价值和应用价值。

但在强调构造性方法的同时，不能忽视非构造性方法，它们是相辅相成的。对此，胡世华曾有过一段十分精

辟的论述。他指出："构造性数学总是自觉或不自觉地在非构造性数学的原则下进行研究和探索；而非构造性数学中又总存在着构造性数学的因素，'纯净'的非构造性数学是不存在的。在非构造性数学的研究中，构造性成分越多的部分往往对自身的发展也越有意义。"[5] 例如，要使计算机为人服务，就必须使数学规律计算化、算法化，就要研究计算数学，研究构造性数学。但是，研究用计算机来做事的可能性、有效性、可行性以至局限性，又非进入非构造性的研究不可。

3. 中国古算的构造性特征

中国古算具有鲜明的构造性特征。在数学起源问题上，有所谓"河出图，洛出书，圣人则之"[12] 的传说。所谓"洛书"，即为今之三阶幻方，它具有美妙的结构：1 至 9 九个自然数构成的纵、横、斜三数之和均为 15。这种幻方，和诸如点、线、面、体之类的几何对象不同，在现实中是找不到它的原型的。虽然，数学起源河图、洛书之说纯属无



洛书图

图 2.4

稽之谈，但作为心灵创造的幻方 —— 洛书（图 2.4）最早出现在中国，确实显示了中国传统数学的构造性特点和

中算家的卓越创造才能。

构造性数学的特点在于它是从无到有的发明，中算家构造的十进位位值制记数法和线性方程组的筹码图阵，就是这种美妙发明之一。例如，《九章算术》卷 8 第一问是："今有上禾三秉、中禾二秉、下禾一秉，实三十九斗；上禾二秉、中禾三秉、下禾一秉，实三十四斗；上禾一秉、中禾二秉、下禾三秉，实二十六斗。问上、中、下禾实一秉各几何？"

图 2.5

其术曰："置上禾三秉、中禾二秉、下禾一秉，实三十九斗于右方，中左禾列如右方。"按《九章算术》的这段术文用筹码数字记录下来为图 2.5。即是现代数学中的矩阵，用以表示"方程"，即今之线性方程组(*)：

$$\begin{cases} 3x + 2y + z = 39, \\ 2x + 3y + z = 34, \\ x + 2y + 3z = 26. \end{cases} \quad (*)$$

在处理不尽方根问题时，刘徽与古希腊学者不同。希腊人在证明了无理数 $\sqrt{2}$ 存在以后，就此止步，而中算家不仅"以面命之"，承认不尽根的合法地位，刘徽更用十进分数无限逼近它，这与近代西方以无穷小数来表示实数，已是一衣带水，表现出强烈的构造性倾向。

在几何上，中算家通过巧妙的构图，将几何量之间的内在联系呈现在人们面前。从而实现了定理、公式的证明，其智慧集中反映在数学对象的构造上。其中，刘徽更有发展。他对“牟合方盖”的发明，球体积公式的发现和刘徽原理的证明等都是这种智慧的杰出表现。

总之，中国古算着眼于问题的求解，算家以设计出尽量简单的程序来求出其解答为目标，一般不考虑解的唯一性等问题。在几何问题的处理上，并不追求逻辑论证的完美，而是凭藉巧妙的构图进行“析理”，以达到论证的目的，这些都充分表现了中国古代传统数学的构造性特色。

4. 幻方与魔阵的构造

作为构造性方法应用之例，我们来考察幻方与魔阵的构造。

一个 n 阶幻方是指将 n^2 个正整数 $1,2,3,\cdots,n^2$ 排成的一个 $n\times n$ 方阵，且使每行、每列、每斜对角线上各数之和均等于同一数 S。S 称为幻方的幻和。易于证明，如果 n 阶幻方存在的话，它的幻和必为

$$S=\frac{n(n^2+1)}{2}。$$

事实上，设有组成 n 阶幻方的 n^2 个正整数，它们成等差数列，a 是首项，公差为 d，则幻和

$$S=na+\frac{(n^2-1)n}{2}d。$$

取 $a=1$、$d=1$ 时，

$$S=\frac{n(n^2+1)}{2}。$$

当 $n=3,4,5,6,7,8$，幻和分别为 15，34，65，111，175，260。

“洛书”作为三阶幻方，古人称为纵横图。其构造方

法，我国宋代杨辉的《续古摘奇算法》(1275) 卷一“纵横图”中记载道：“洛书：九子斜排，上下对易，左右相更，四维挺出，载九履一，左三右七，二四为肩，六八为足。”也即首先把 1,2,3,…,9 斜排成(1) 的斜直线三排，然后上下对易，左右相更，便得洛书图。

		1		
	4		2	
7		5		3
	8		6	
		9		

(1)

4	9	2
3	5	7
8	1	6

(2)

这个方法可以推广至任意的奇阶魔方。若 n 是奇数，则可把 $1,2,\cdots,n^2$ 这些数斜排成 n 列，然后把在中间的 $n\times n$ 阶方阵以外的部分进行“上下对易，左右相更”，这种对易和相更是把正方形外的数分别平移到正方形内的相对位置上。以制作五阶幻方为例，先作 5×5 格的正方形和每边均外凸的四个格子，如(1)。自最上格起将 1 至 25 分五排沿斜线填在方格内，再把外凸的数字移到对面内侧的空格中，这样就构成了五阶幻方(2)。

				1				
			6		2			
		11		7		3		
	16		12		8		4	
21		17		13		9		5
	22		18		14		10	
		23		19		15		
			24		20			
				25				

(1)

11	24	7	20	3
4	12	25	8	16
17	5	13	21	9
10	18	1	14	22
23	6	19	2	15

(2)

上述构造方法成功的道理，可作如下说明：

如(2)的第1行中的3，在(1)中是第一斜排中第三斜格的数，记以$C_{13}=3$。同样$C_{22}=7$，$C_{31}=11$，$C_{45}=20$，$C_{54}=24$，而$3=0\times5+3$，$7=1\times5+2$，$11=2\times5+1$，$20=3\times5+5$，$24=4\times5+4$，则第一行诸数之和为

$$(0+1+2+3+4)\times5+(1+2+3+4+5)=65。$$

其他各行、各列及各对角线情况均类似。

杨辉还给出了四阶幻方的构造方法，即把$1,2,\cdots,4^2$顺次直排4列，然后将两对角线上的数字，以方形的中心为对称中心，互换对称位置上的数字，这就构成了4阶幻方。

1	5	9	13
2	6	10	14
3	7	11	15
4	8	12	16

→

16	5	9	4
2	11	7	14
3	10	6	15
13	8	12	1

需要指出的是，2阶幻方是不存在的，3阶幻方只有一个，4阶幻方一定存在，但不唯一。

在杨辉的著作里，共收集了二十多个幻方，阶数n为$3,4,5,\cdots,10$。其中，10阶幻方，杨辉称为“百子图”，即是“纵横五百五，共积五千五十”。还有一种“攒九图”，它是由1到33的数字排成同心圆，9位于中心，且33个数字分布在四条直径上，形成“斜直周围各一百四十七”的形式(图2.6)。

所谓魔阵，是这样一个方形数字图，其形中填有n^2($n\in N$)个数字，且满足如下性质：

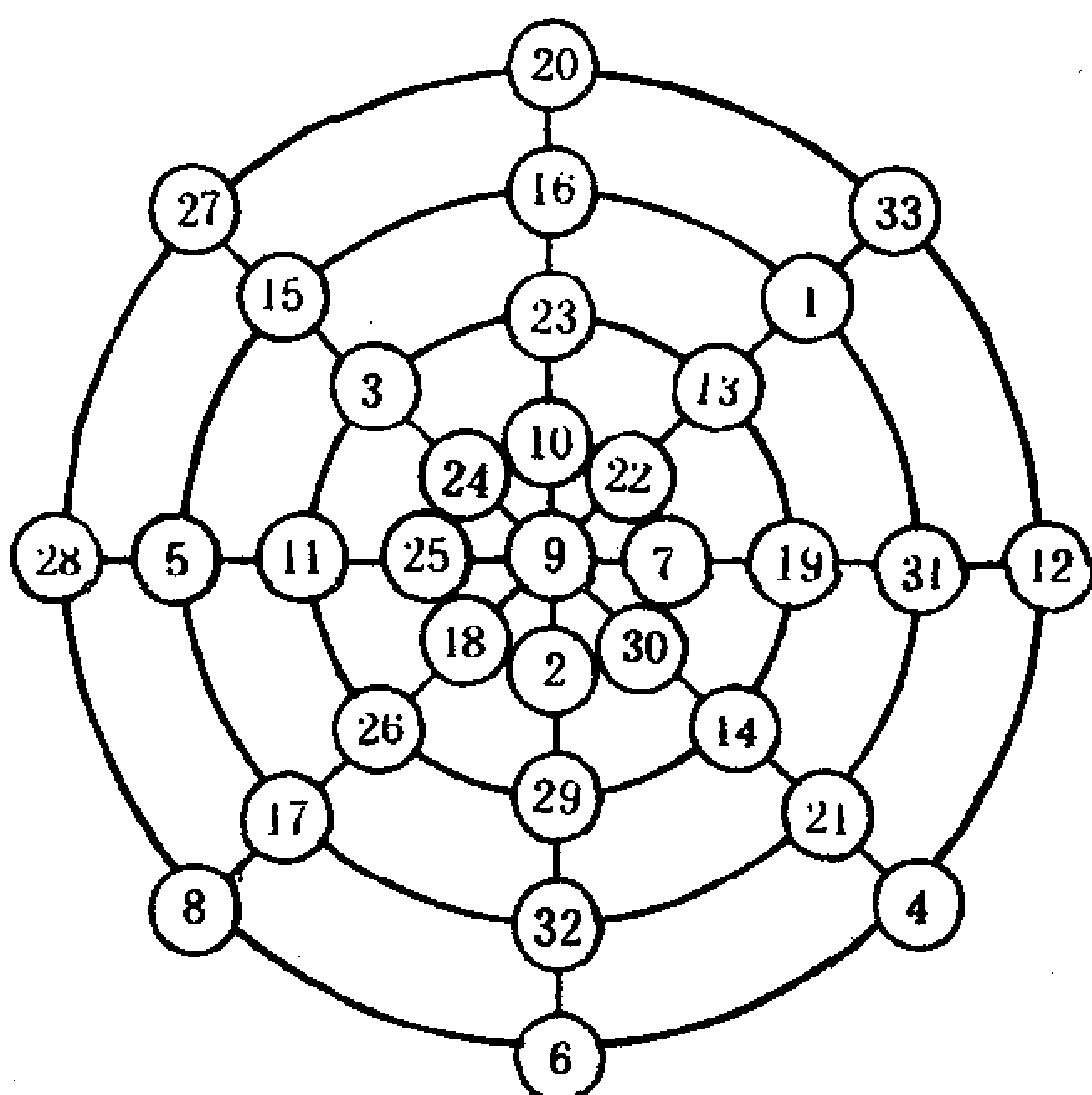


图 2.6

+	8	3	7	2
6	14	9	13	8
5	13	8	12	7
4	12	7	11	6
1	9	4	8	3

任选其中一数字，并将这一数字的同行、同列的数字划去，随后又在余下的数字中任选第二个数字，并划去其同行、同列的数字，如此继续下去，直至选到最后一个数字为止。这样所有选出的数字之和都是一确定的数，这个确定的数称为魔阵的“魔数”。如左图是个魔阵，其魔数为 36。它的构成方式是由横行 8，3，7，2 与竖行 6，5，4，1 两两相加后填入方格构成的。故其魔数必为这八个数字的和即 36。

一般地，设 4×4 魔阵按下表方法构成：

+	a	b	c	d
e	$a+e$	$b+e$	$c+e$	$d+e$
f	$a+f$	$b+f$	$c+f$	$d+f$
g	$a+g$	$b+g$	$c+g$	$d+g$
h	$a+h$	$b+h$	$c+h$	$d+h$

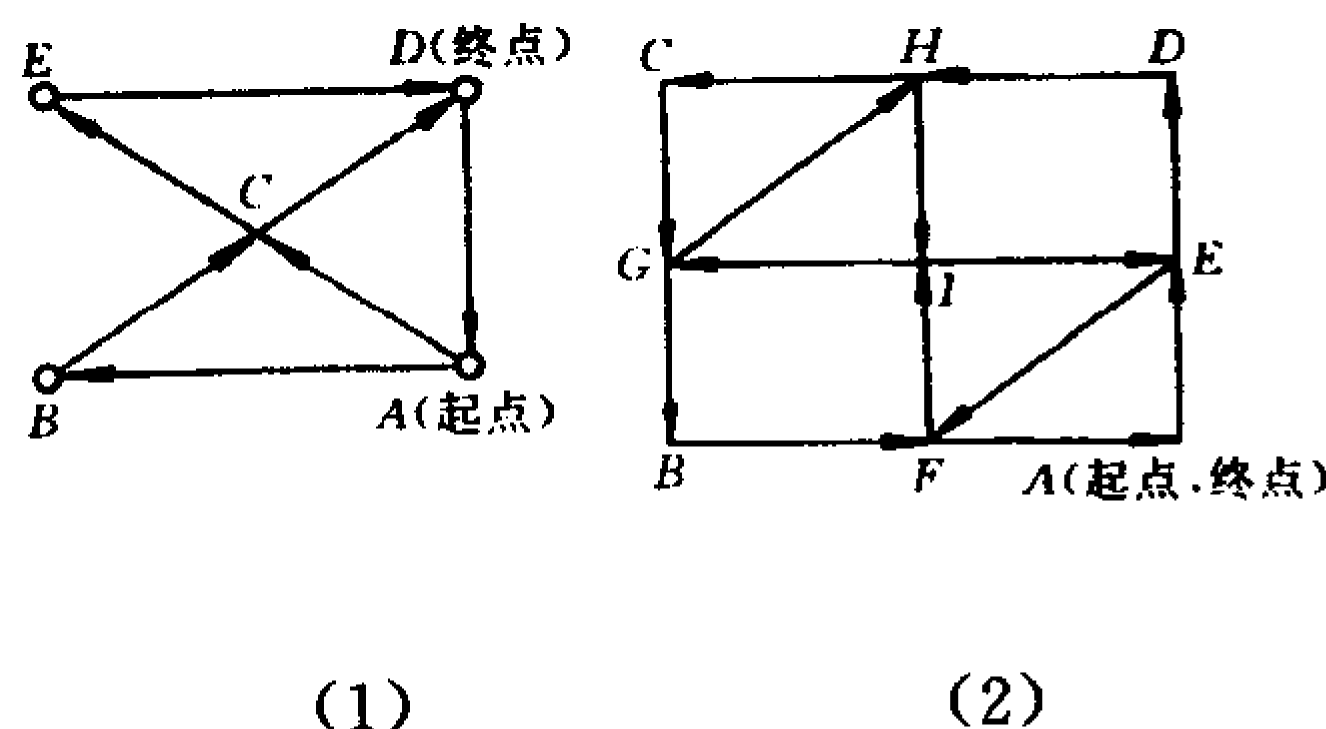
这样，按任一选取和划去的手续，所选数字之和即魔数必为 $a+b+c+d+e+f+g+h$。

5. 一笔画图的构造

一笔画图，是指这样的一种图形，它由若干个点及这些点的一些连线组成，并能够用笔不重复地、不遗漏地一笔画出。

例如，图 2.7 中两个图形就是一笔画图。

这是因为，图(1) 可按 $A\to B\to C\to D\to A\to C\to E\to D$ 的次序一笔画成，而图(2) 也可按 $A\to E\to F\to I\to E\to D\to H\to I\to G\to H\to C\to G\to B\to F\to A$ 的次序一笔画成。



(1)　　(2)

图 2.7

若将点按与它相连的线的数目的奇偶数区分，它可分成奇点和偶点。这时，左图中奇点为 A、D，其个数为 2，且这两个奇点分别是一笔画图的起点和终点。而右图的所有点都是偶点，这时一笔画图的起点和终点重合。

对于一笔画图，有这样的判定准则，且这一判定准则是充分必要的。一个连通的图，若它只有 0 个或 2 个奇点，则该图一定可以一笔画成，且若奇点个数为 0 时，则可任找一个点作为起点和终点(如图 2.7(2))。若奇点个数为 2 时，则这两个奇点可以分别作为起点和终点(如图 2.7(1))。

实际中，一些最短线路问题，可归结为一笔画图的构造问题。

例如，有某风景区游览图(图 2.8)，线上的数字表示路线距离的公里数。现要从车站出发，每个地方、每条道路都参观游览一次，然后回到车站，试安排一条最短的参观游览路线。

易知，图中除庙、湖两个奇点外，其余都是偶点，这是一笔画图。但这种一笔画图要分别以两个奇点作为起点和终点。按题意，游览需从车站出发回到车站，故需将上图改为奇点数为 0 的图，亦即应使庙与湖都改为偶点，因此要设法在这两点之间增加一条连线，这相当于重复多走一段路。现从庙到湖的各路程中，以庙 — 亭 — 湖为最短，故应选择重复多走这段路。于是，符合要求的最短参观游览线路可以为：

车站 → 博物馆 → 寺 → 亭 → 博物馆 → 庙 → 亭 → 湖 → 亭 → 庙 → 湖 → 寺 → 车站。

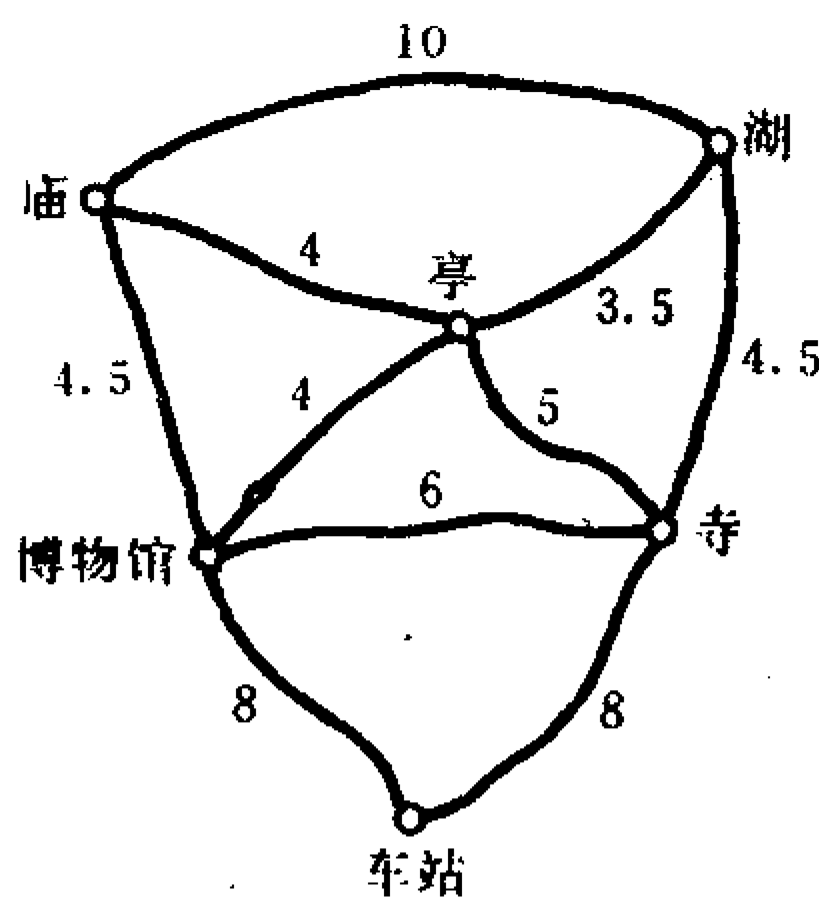


图 2.8

构造一笔画图在实际中应用很广，如纺织工人在车间的巡回线路，邮递员投递信件线路等都可归结为按实际问题需要构造一笔画图问题。

2.4　数学模型思维方式

数学模型思维方式是数学刻画客观事物原型并用以研究、解决实际问题的工具，它也是数学科学联结其他非数学科学的中介和桥梁。

1. 数学模型及其构建方法

模型一词，英文称“model”，包含有“模型”、“维形”、“模式”的意思。

数学模型，是指对某种事物系统的特征和数量关系，借助数学语言而建立起来的符号系统。数学模型有广义和狭义两种。广义是指凡以相应的客观原型作为背景加以一级或多级抽象的数学概念、理论、式子等，都称为数学模型。狭义是专指反映某特定问题或事物的数学符号系统。在应用数学中，数学模型的含义通常按狭义来理解，其目的在于解决具体、实际的问题。

数学模型可按不同的标准来区分。例如，就人们对其认识过程而言，可区分为描述性模型和解释性模型。描述性模型是从特殊到一般，它是从分析具体事物的量的关系出发，通过抽象将这种关系概括在一个数学结构之中的一种模型。而解释性模型，则是由一般到特殊，即是从一般的公理系统出发，借助于数学客体，来对公理系统作出解释的一种模型。在描述性数学模型中，又可按数学研究的实际对象区分出不同类型的模型。例如，微积分是运动过程的数学模型，是一种确定性模型；概率统计是随机现象的数学模型，是一种或然性模型。

数学模型构造的一般过程为：建模准备 —— 模型的假设和建立 —— 模型求解 —— 模型检验。其中准备阶

段的工作主要包括了解现实对象的实际背景，明确建模的目的、要求，收集有关数据，做好各种准备等。下面以描述性模型为例来讨论一个具体实际问题建模过程的后三步。

设两个人与两只熊要过河，现有一只空船能自动来回运驶。已知船最多能载两个乘客，又若人比熊少时，熊会吃掉人，若每只熊能划船，问要过河需行船几次？且有多少种不同的过河方法？

建模过程：

(1) 提炼数学模型。由于行船的最后一次总是只过而不回，故行船的次数必为奇数。

若把 m 个人和 n 只熊在同一边的状态记作 (m,n)，则问题的情况共有 $9(=3^2)$ 种：

$(2,2)$，$(0,0)$；$(0,2)$，$(2,0)$；$(2,1)$，$(0,1)$；$(1,2)$，$(1,0)$；$(1,1)$。

若把同一时刻河两岸所处的状态 (m,n) 和 (m',n') 称为互补的，其中

$$m+m'=2,n+n'=2。$$

前四组中每两个状态是互补的，最后 $(1,1)$ 与自己互补。注意到 $(1,2)$ 与 $(1,0)$ 应剔除，因为在 $(1,2)$ 时熊比人多，而 $(1,0)$ 是 $(1,2)$ 的互补状态。

若除去不出现的状态，且互补状态中只取其一，故只剩下四种，记

$$\mathrm{I}=(2,2),\mathrm{II}=(0,2),\mathrm{III}=(2,1),\mathrm{IV}=(1,1)。$$

则问题就变更为使状态 I 转移为 IV 且转移次数为奇数时，有多少种不同方法？

设用矩阵表示状态的转移。先作出表示一次转移的矩阵 V^1，某元素 a_{ij} 表示由状态 i 经过一次转移成为状态

j 的所有可能走法的种数。例如，状态 Ⅲ 表示两个人、一只熊，这时对岸仅有一只熊，若让另一只熊乘船过河，对岸便出现状态 Ⅱ，故 $a_{32}=1$，而从 Ⅲ 到 Ⅰ 不能经过一次转移实现，故 $a_{31}=0$。仿此可写出矩阵

$$V^1=\begin{bmatrix}0&1&0&1\\1&0&1&0\\0&1&1&1\\1&0&1&0\end{bmatrix}。$$

这样，原实际问题就化为了下面的数学问题：

V^1 经过 k 次（奇数次）变换后，左上角元素为非零，这最小的奇数 k 即为所求的划船次数。

(2) 求解数学模型。经过若干次转移后状态变化的问题，实际是一个若干次一次转移矩阵的乘积问题。计算

$$V^3=V^1\cdot V^1\cdot V^1=\begin{bmatrix}0&4&2&4\\4&1&5&1\\2&5&5&5\\4&1&5&1\end{bmatrix},$$

$$V^4=V^3\cdot V^1=\begin{bmatrix}8&2&10&2\\2&9&7&9\\10&7&15&7\\2&9&7&9\end{bmatrix},$$

$$V^5=V^4\cdot V^1=\begin{bmatrix}4&18&14&18\\18&9&25&9\\14&25&29&25\\18&9&25&9\end{bmatrix}。$$

因 V^5 左上角元素为非零元素 4，故人和熊全部过河最少要经过 5 次转移，并有 4 种不同走法。

(3) 检验数学解答。若记人为"○"，熊为"⊗"，通过

试验知有 4 种过河方法，每种方法需 5 次转移，如图 2-9 所示。

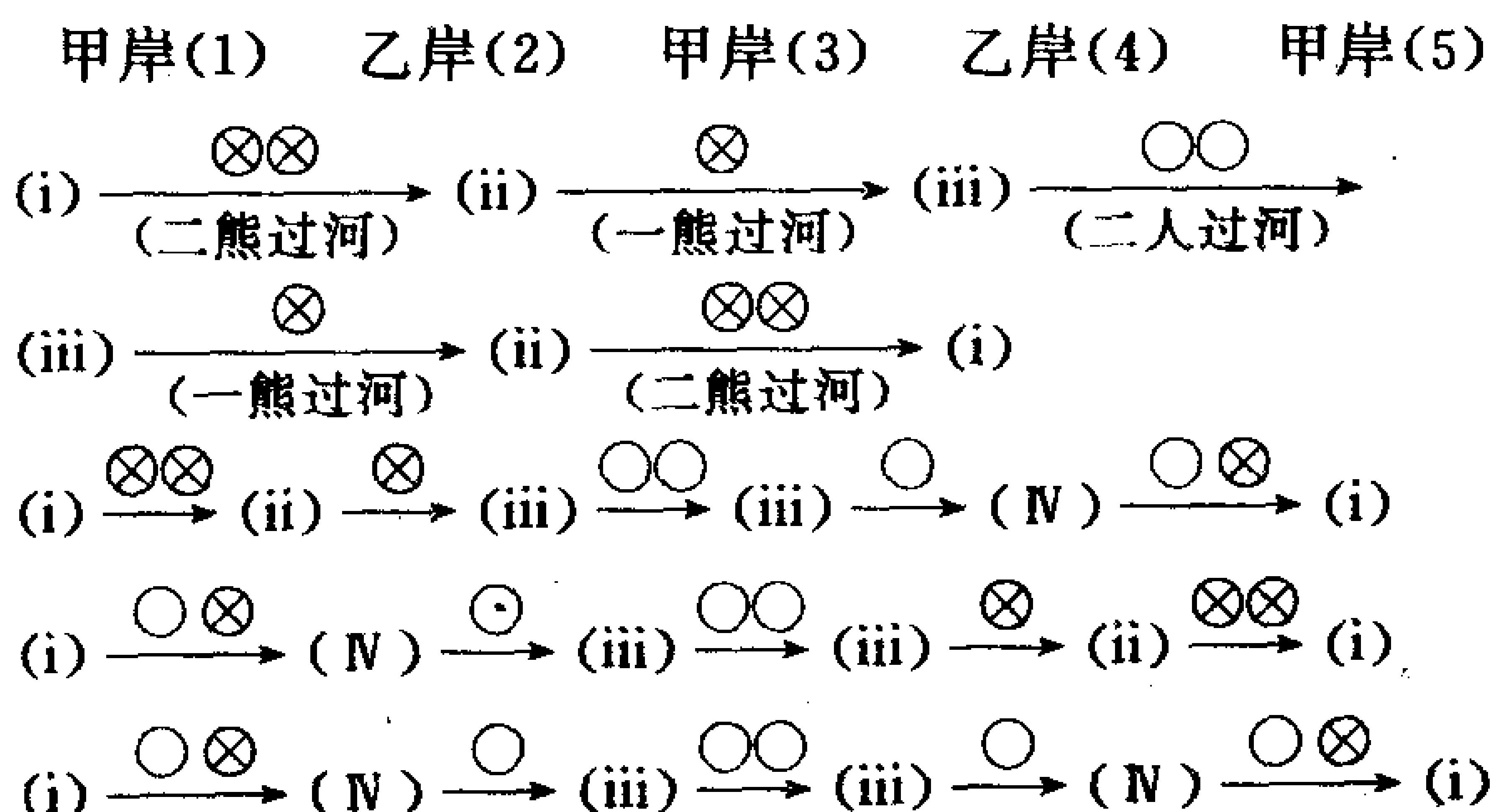


图 2.9

2. 数学模型的功能

数学模型就其功能而言，主要表现为：

(1) 简化性功能。

数学模型从客观事物和现象中抽象出来用以代替现实原型，已经舍去了非本质现象和偶然现象，抓住了主要矛盾和本质特征。因而，数学模型方法具有简化性功能，它是科学研究中简单原理的再现。

(2) 近似性功能。

数学模型的建立突出了实际研究对象的主要特征，忽略了次要特征。所以，这是一种近似性的研究。但由于它抓住了事物的主要矛盾，并以科学抽象为基础，故其研究结果必然基本符合客观实际。且人们对数学模型在此基础上的逐次修正，必然使它越来越逼近于真实实际，这就是它的近似性功能。

(3) 解释性功能。

解释性模型可使抽象的公理系统变得易于理解，并借助它反过来对公理系统中的许多问题，如相容性、独立

性、完备性等进行分析、思考和论证，故它具有解释性功能。

(4) 应用性功能。

对于客观世界存在的尚未被人们认识的规律，通过数学模型方法的探索被揭示出来后，它必然反过来作用于客观世界而获得应用，这就是它的应用性功能。同时，由于数学模型方法实现了对问题陈述、过程刻画和定量计算的符号化，因而大大简化了思维的过程，加速了思维的进程。故这种应用往往非同一般，而常表现为对科学的预见和创新。

例如，19 世纪初，意大利天文学家皮亚齐用望远镜发现了一颗名为谷神星的小行星，但这一行星在接近太阳时消失了，这一现象引起了人们对谷神星的怀疑。不久，数学家高斯用数学计算方法推算出了行星的位置，沿着高斯指出的方向观察，天文学家不仅又找到了这颗行星，而且还先后发现了许多新的行星。

现代社会中，数学模型常可为科学决策提供依据。一个典型的例子是关于核武器发展方向的决策问题。发展核武器要提高核武器的毁伤力，其中，最为重要的两个因素是炸弹的爆炸力和命中精度。对此，美国人建立了一个数学模型

$$k = y^{\frac{2}{3}}/c^2,$$

其中 k 为毁伤力值，y 为炸弹的 T.N.T 当量，c 为命中精度。按此模型，当 y 提高 8 倍时，k 值提高 4 倍；当 c 值减少到原来的 $\frac{1}{8}$ 时，k 值将提高 64 倍。故发展核武器的方向要以提高命中精度为主，在实践上也的确如此。故这一数学模型对科学决策起了重要作用。由于人们认识世界的

全部目的在于改造世界，故数学模型的应用性功能在它的诸功能中占有主导地位。其意义，如同徐利治教授所指出的："现代各门应用数学所以具有解决实际问题的功能，主要就是通过提供数学模型方法而显示出来的。"[20]

3. 投票模型

投票模型的产生除西方国家实行多党制带来的若干政党联合执政问题这一背景外，在经济联合委员会决策、股东权益等方面也可找到它的原型。

设有 n 个政党，$w_i(i=1,2,\cdots,n)$ 为第 i 个政党在国会中拥有的席位，则全部议席数为

$$W=\sum_{i=1}^{n}w_i。$$

若以 q(定值) 表示执政所需的最低议席数，则 $[q;w_1,w_2,\cdots,w_n]$ 可看作是 n 个政党分别带权 $w_1,\cdots,w_n$ 的一次竞争。如果有一集合 S，它是 $N=\{1,2,\cdots,n\}$ 的一个子集，使得

$$\sum_{i\in S}w_i\geqslant q,$$

则 S 上各政党的联合取得多数席时可以执政，q 通常取大于 $\frac{W}{2}$，且

$$w_i<q(i=1,2,\cdots,n)。$$

例如，1972 年加拿大选举，自由党(a)、保守党(b)、新民主党(c) 分别得 109、107、31 席，其他力量(d) 占 17 席，总席位是 264，超过半数为多数。由于上述三党任何二方联合均可执政，故他们席位虽不同，但处于同等有利的地位。不过这样的分析仅依赖于直观，要进行数量刻画。

设竞赛格局是 $[q;w_1,\cdots,w_n]$，S 是集合 $N=\{1,2,$

$\cdots n\}$ 的子集，$W = \{S \mid \sum_{i\in S} w_i \geq q\}$ 表示所有可赢的联合，M 表示全体最小的可赢联合，即少了 S 中的任何一个都不行。$M(i)$ 表示有 i 参加的最小可赢联合体。通常定义各竞争者的实力指标为

$$\rho(i) = \frac{1}{|M|}\sum_{S\in M(i)} \frac{1}{|S|}。$$

其中，$|M|$、$|S|$ 分别表示集合 M、S 元素的个数。

设有竞赛

$\Gamma = \{5;4,2,1,1,1\}$，

竞争者依次为 a,b,c,d,e，则各竞争者实力指标的求法为：

全体最小可赢联合 $M = \{ab,ac,ad,ae,bcde\}$，

a 参加的最小可赢联合 $M(a) = \{ab,ac,ad,ae\}$，

$$\rho(a) = \frac{1}{5}\left(\frac{1}{2}+\frac{1}{2}+\frac{1}{2}+\frac{1}{2}\right) = \frac{8}{20}。$$

而

$M(b) = \{ab,bcde\}$，

故　$\rho(b) = \frac{1}{5}\left(\frac{1}{2}+\frac{1}{4}\right) = \frac{3}{20}$。

又显然 $\rho(c) = \rho(b) = \rho(d) = \rho(e)$，因此，上述 1972 年加拿大选举问题，归结为：

竞赛 $\Gamma = \{132;109,107,31,17\}$ 且竞争者为 a、b、c、d。

由于 $M = \{ab,ac,bc\}$，故 $\rho(a) = \rho(b) = \rho(c) = \frac{1}{3}\left(\frac{1}{2}+\frac{1}{2}\right) = \frac{1}{3}$，其结果与直观分析相一致。

以上关于多党制联合执政的刻画同样适用于公司中 n 位股东的决策。设有 4 个股东 A、B、C、D，所持的股票份额分别为 28，24，24，24。这时，四个股东的竞争实力便归

结为竞赛 $T=\{51;28,24,24,24\}$ 中 $\rho(A)$、$\rho(B)$、$\rho(C)$、$\rho(D)$ 的计算问题。

4. 信息量模型

信息作为科学概念是 1948 年美国贝尔电话公司的申农(C. E. Shannon) 博士首先提出的。申农认为,信息乃是“两次不定性之差”。“不定性”亦即人们对客观事物的不了解、不确定,当人们获得新知识后,改变了原有的知识状态,便减少或消除了原先的“不定性”。他把这种不定性减少或消除的数量称为信息量,故信息量是收信人知识变化的数值度量。据此,申农建立了信息量统计的数学模型。

设用 N 个字母通信,字母集为 $A=\{\alpha_1,\alpha_2,\cdots,\alpha_N\}$。$p_1,p_2,\cdots,p_N$ 分别为上述字母在通信中使用的概率。$p_i\geqslant 0,\sum\limits_{i=1}^{N}p_i=1$,

此时,每一字母 α_i 所产生的信息量为 $-\log_2 p_i$,称为自信息。将自信息作加权平均,即得

$$H(A)=-\sum_{i=1}^{N}p_i\log_2 p_i,$$

$H(A)$ 称为 A 的信息量,这里的单位叫做二进单位,或称作比特(bit)。其意义是:一比特的信息量,就是含有两个独立等概率可能状态的事件所具有的不确定性被全部消除所需要的信息。信息的单位选择,取决于所选择对数的底。取自然对数,信息单位为奈特,取常用对数,信息单位为哈特。

例如,我国古代的烽火台,燃起烽火表示有敌情,不燃则无敌情。这时共有两种信号,其信息量为 1 比特。

又如,某信源,它产生一个由 A、B、C、D 中选出的字

母序列，概率分别为$\frac{1}{2}$，$\frac{1}{4}$，$\frac{1}{8}$，$\frac{1}{8}$，并且相继符号的选择是独立的。则

$$H = -(2^{-1}\log_2 2^{-1} + 2^{-2}\log_2 2^{-2} + 2 \times 2^{-3}\log_2 2^{-3})$$

$$= \frac{7}{4}(\text{比特 / 符号})$$

值得注意的是，上述信息量的数学表达式和热力学中的熵的公式在形式上完全相似，仅差一个符号。故人们把信息称为负熵，从而导致了对信息概念的更深刻的理解。在热力学中，熵表示系统的紊乱程度，即在一个孤立的系统中，总是自发地从有序到无序，它是系统不确定性的度量。对负熵，信息量愈大，信息越多，不确定性越小，无序情形则越少，故信息量是系统有序状态的度量，它标志着系统的确定性程度。信息作为一切组织系统的一种普遍联系形式，其主要变化标志着组织系统的兴衰。故它的数学模型为信息的广泛应用奠定了基础。

5. 化学方程式平衡模型

试平衡化学方程式：



$?HNO_3 + ?Cu = ?Cu(NO_3)_2 + ?H_2O + ?NO$。

问题是求方程中各分子式的系数，设它们顺次为z_1，z_2，和$-z_3$，$-z_4$，$-z_5$。则原问题化为从

$$z_1 HNO_3 + z_2 Cu + z_3 Cu(NO_3)_2 + z_4 H_2O + z_5 NO = 0$$

中求z_1, z_2, z_3, z_4, z_5。若以元素H、N、O、Cu为顺序，则方程中各反应物、生成物相应对应一有限维非零向量：

$HNO_3 \leftrightarrow [1,1,3,0]^T$，$Cu \leftrightarrow [0,0,0,1]^T$，

$Cu(NO_3)_2 \leftrightarrow [0,2,6,1]^T$，$H_2O \leftrightarrow [2,0,1,0]^T$，

$NO \leftrightarrow [0,1,1,0]^T$。

这样问题变为从

$$
\begin{array}{c} \\ \text{H} \\ \text{N} \\ \text{O} \\ \text{Cu} \end{array}
\begin{array}{c}
\begin{array}{ccccc} HNO_3 & Cu & Cu(NO_3)_2 & H_2O & NO \end{array} \\
\begin{bmatrix} 1 & 0 & 0 & 2 & 0 \\ 1 & 0 & 2 & 0 & 1 \\ 3 & 0 & 6 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 \end{bmatrix}
\end{array}
\begin{bmatrix} z_1 \\ \vdots \\ z_5 \end{bmatrix} = 0 \qquad (*)
$$

中求$[z_1,z_2,\cdots,z_5]^T$。

令

$$[z_1,z_2,\cdots,z_5]=Z,$$

(*)中与Z^T相乘的矩阵为C,这时C是4×5的矩阵,4是化学元素的个数,5是反应物与生成物个数之和。

一般地,设方程中共出现n种元素,i种反应物,j种生成物($i+j=m$),则化学方程式的平衡归结为求系数向量$Z=[z_1,z_2,\cdots,z_m]$,满足$CZ^T=0$的适合问题的最小解,其中Z的分量为非零有理数,且当分量为正时代表反应物,为负时代表生成物,而C是这样一个$n\times m$矩阵,其生成方式是:

将元素按顺序编号为$1,2,\cdots,n$,反应物和生成物顺次编号为$1,2,\cdots,m$,每种反应物、生成物对应着按元素顺序排列的一个n维非零列向量,记为$[C_{k1},C_{k2},\cdots,C_{kn}]^T(k=1,2,\cdots,m)$。则$C$是由形如$[C_{k1},C_{k2},\cdots,C_{kn}]^T$的$m$个列向量组成的$n\times m$矩阵,它的元均为非负整数。

这便是化学方程式平衡的数学模型。简言之,它归结为一个求线性方程组的适合题意的最小解。上例中,对(*)求解,得

$$[z_1,z_2,\cdots,z_5]^T=k[8,3,-3,-4,-2]^T。$$

作为方程式系数,取$k=1$,则有



$$8HNO_3+3Cu=3Cu(NO_3)_2+4H_2O+2NO。$$

2.5　公理化思维方式

公理化思维方式是一种重要的数学思维方式。这种思维方式不仅对数学，而且对其他自然科学都有着极为重要的作用。

1. 公理和公理化方法

数学的公理化方法是从尽可能少的基本概念和公理出发，应用严格的逻辑推演，把数学的某分支组织成为演绎系统的一种方法。在这样的演绎系统中，基本概念是无定义的或隐定义(其涵义由公理来确定)的原始概念。而所谓公理则是无条件承认的相互制约的规定，亦即"是数学需要用作自己的出发点的少数思想上的规定"[6]。

公理化方法在数学史上经历了实质公理化方法——形式公理化方法——元数学研究的过程，且这种发展集中反映了数学抽象层次的不断提高上。

(1) 实质公理化方法。

实质公理化方法是指"对象——公理——演绎"系统，其特征是公理的对象是从属于某种特定对象。欧几里得(Euclid)的《几何原本》是这种方法的典型代表。即是说，这些对象是以现实空间作为直观背景的。如在《几何原本》中，欧几里得明确规定了研究的对象是"没有部分的"("点")，"没有宽度的"("线")和"只有宽度和长度的"("面")，他用作推理依据的公理、公设都是这些对象的明显事实。这种从属于特定的现实空间的公理系统就是实质公理系统。实质公理系统反映了当时数学发展的水平：即在很长时间内数学的研究仅限于算术与几何这种具有明确直观背景的内容，故公理体系也就被认为这

些对象所特定的，以致很长时间人们一直把欧氏几何当做关于空间的绝对真理，而把算术当做数量关系的唯一理论。

(2) 形式公理化方法。

形式公理化方法即“假设 —— 演绎”系统。与实质公理系统相比，在这样的公理系统中，公理不再具有“自明性”或“必然性”，而只是作为演绎基础的“假设”，且不再是由对象决定公理，而是以公理来决定对象。即对象可以是任何东西，只要能满足系统中诸公理的要求。形式公理系统的典型代表是希尔伯特的《几何基础》(1899 年出版)。希尔伯特对传统的几何题材进行了新的处理，他建立的演绎系统的推理出发点是五组公理。“在 … 之上”、“在 … 之间”、“合同于”、“平行于”、“连续”等为基本概念，概念的严格含义是由公理决定的，而对象也由公理决定，只要满足给定的公理，称它们为什么是无关紧要的。正如希尔伯特所说的：“我们必定可以用桌子、椅子和啤酒杯来代替点、线、面。”正是由于形式公理方法反映的不只是特定的研究对象的性质，而是许多具有相同结构的对象的共同性质，故与实质公理方法相比，它对数学具有更为重要的意义。

(3) 元数学。

元数学是把某种数学理论作为整体来加以研究。与前两阶段相比，如果说以前是在系统内部进行研究的话，现在则是对系统本身加以研究，它标志着公理化方法发展到了一个更新的阶段。希尔伯特关于建立元数学的基本思想是这样的：他认为，除了我们前面所说的数学理论的公理化和形式化之外，我们必须着眼于把整个系统作为研究对象，这便大大有别于以往的研究方法，有别于以

往的着眼点。通常称被研究的数学理论为对象理论，而用以研究或作为研究工具的那个理论叫做元理论。所获定理，若为有关整个理论系统之某种性质的定理叫做元定理。例如，我们想证明“整个数学理论是相容的”，这就是一条元定理。它是着眼于整个数学理论而言的结论，不同于数学系统内的任何一条具体定理。

2. 公理化方法的作用

(1) 公理化方法是总结和整理数学知识并使之系统化、科学化的有效方法。

人们在实践中积累了丰富资料和大量知识后，要按照事物内部的逻辑联系进行“去粗取精、去伪存真、由此及彼、由表及里”的加工制作，以造成系统的理论，公理化方法正是建立这种理论的有效方法。因为经过公理化方法处理后的知识，已经不是零散的感性知识的堆积，而是按照对象的内在联系逻辑地沟通的，故公理化方法是建立数学理论不可缺少的方法。例如，公元前 300 年左右，欧几里得继承前人研究的成果，将人们积累的丰富几何知识，用公理方法加以整理，使之条理化、系统化，从而在数学史上第一次建立了较为系统、严密的几何理论。这种用公理方法整理组织数学理论的方法带来了深刻的影响。尔后的数学家都用这一方法建立数学理论，出现了诸如群论、概率论、公理集合论等近代数学的公理系统。

(2) 公理化方法是检验数学命题正确性的重要手段。

某一数学分支经过公理化后成为一个严密的演绎系统，这时任一数学命题的真假都可通过逻辑来判断。对此，M. 克莱因曾论述道：“数学主要是一种寻求众所周知的公理法思想的方法。这种方法包括明确地表述出将要

讨论的概念的定义，以及准确地表述出作为推理基础的公设，具有极其严密的逻辑思维能力的人从这些定义和公设出发，推导出结论。数学的这一特征由17世纪一位著名的作家在论及数学和科学时，以某种不同的方式表述过：'数学家们像恋人…… 承认一位数学家的最初的原理，那么他由此将会推导出你也必须承认的另一结论，从这一结论又推导出其他的结论。'"[5] 因此，公理化方法是检验数学命题正确性的重要手段。当然，通过公理化方法检验而获得的命题的真实性仅具有模式真理的性质，它不等同于现实真理性。其中，前者是相对于数学模式借助于逻辑定义而获得的稳定的关系结构而言的，后者则是指数学模式所具有的现实意义，即是指它们反映了真实世界中的某种关系形式或特征。这样的两种真理性的产生，是由于以下的事实："现代理论数学的研究对象已不仅是已经给出的(包括具有明显现实意义的)数学构造，而且还包括了理论上各种可能的数学构造，甚至是现实经验中未能找到的构造。"[5]

(3) 公理化方法是推动和促进数学理论创新的一种活力。

公理化方法通过探索事物发展的逻辑规律，考察它们之间的逻辑联系，因而易于从逻辑上发现问题、提出问题，这往往是某种新数学理论创立的开始。数学史上，鉴于欧几里得几何体系的重大影响，数学家们很早就对欧几里得公理系统进行研究，希望把欧氏公理体系中的公设减少到最低限度，亦即试图从其他公理中导出第五公设，这种纯逻辑的探讨导致了非欧几何的产生。公理化方法对数学理论创立的推动，在抽象代数中亦表现得非常突出，完全如著名代数学家范德·瓦尔登所说："近世代

数的扩大主要是由于公理方法，使用这个方法产生了一系列新的概念…… 而且得到了许多有深远意义的成果，特别是域论、理想数论、群论和结合代数方面。”[10] 尤其是公理化方法从实质的进入到形式的阶段后，它所反映的已不只是某种特定的研究对象的性质，而是许多具有相同数学结构的对象的共同性质，从而对数学理论的创新具有更重要的意义。

(4) 公理化方法是研究自然科学的重要方法。

数学是表达自然科学的语言，而公理化方法则是这种语言的精粹，因为它最简明、最严格。因此，公理化方法对于自然科学，也是不可缺少的。科学史上，牛顿的《自然哲学的数学原理》(1686)、拉格朗日的《解析力学》、克劳修斯的《热的机械运动理论》等奠基性论著，都是在效仿欧几里得公理化方法的思想指导下完成的。尤为突出的是，公理化方法在近代自然科学理论发展中也屡建功勋。例如，在建立广义相对论工作中，希尔伯特充分发挥了数学家的优势，他用变分法、不变式论等强有力的数学工具，按公理方法进行研究，取得了决定性的成果，他在一份报告中概述了这项工作。他说：“遵循公理化方法，事实上从两条简单的公理出发，我要提出一组新的物理学基本方程，这组方程具有漂亮的理想形式，并且我相信它们同时包含了爱因斯坦与 G. 米所提问题的解答。”[10]

(5) 公理化的思想方法是人工智能系统建立的不可缺少的条件。

公理化方法的深入发展，使之更加形式化、精确化、符号化，从而使电子计算机模拟人脑思维成为可能。数学史上正是由于对公理化方法和数学基础的深入研究，才促使了逻辑实现了从普通的走向符号的进步而产生了数

理逻辑，从而促进了人工智能的发展。没有公理化思想，就不可能有人工智能的开发和应用，故它是人工智能系统建立的不可缺少的条件。

3. 公理体系的三个基本问题

一个科学的公理体系，不是一些基本概念和公理的任意罗列，而必须满足一定的要求，这就是公理体系的相容性、独立性和完备性。

(1) 相容性。

相容性即和谐性，无矛盾性。其意是，在给定的公理体系内不论推论到多远，决不允许出现任何相互矛盾的命题。这是构成公理体系的首要条件。

关于一个公理系统的相容性，有所谓相对相容和直接相容之分。其中，所谓相对相容，指的是依靠某一个公理系统的相容性为前提，去证明另一个公理系统的相容性，此时就称所说的另一个公理系统获得了相对相容性证明。而当不依靠任何别的公理系统的相容性为前提，而直接证明了一个公理系统的相容性，就叫做给出了该公理系统的直接的相容性证明。而关于公理系统的相对相容性证明往往采用模型方法。例如，庞加莱在欧氏几何中构造罗氏几何的模型。又笛卡儿创立的解析几何启示我们能在实数系统中构造欧氏几何模型。所以在实数系统被假定为相容的前提下，欧氏几何与罗氏几何都是相容的，但都是相对相容而已。

(2) 独立性。

独立性是指公理体系里的每一条公理都是必要的。即其中的每一条公理都不能从其他公理中推出。故独立性问题是保留同样多推论的前提下，公理体系所含公理最少个数的问题。

设公理体系 Σ 包含有 n 条公理，记作 $\Sigma = \{A_1, \cdots, A_n\}$，又 $\overline{A}$ 表示 A 的矛盾命题，“公理体系 Σ 不能推出某一命题 P”记以“$\Sigma \nRightarrow P$”，则证明 A_i 在 Σ 中是独立的，就是要寻找一个模型 M，在其中 $A_1, A_2, \cdots, A_{i-1}, A_{i+1}, \cdots, A_n$ 都成立，而 A_i 却不成立，亦即

$$\Sigma' = \{A_1, A_2, \cdots, A_{i-1}, \overline{A}_i, A_{i+1}, \cdots, A_n\}$$

相容。

因此，若公理 Σ 中有 n 条公理，要证明 Σ 的独立性，就要构造 n 个模型。

例如，若将希尔伯特公理体系中除平行公理 V 外的其余公理称为绝对几何公理体系，记以 Σ_0。由于 Σ_0 存在罗氏几何模型使 $\Sigma_0 + \overline{V}$ 为相容，故 V 是独立的。

(3) 完备性。

一个公理体系是完备的是指，若不能再增加新的公理到该体系中使之成为一个更细致的公理体系。亦即，在此公理体系中能确保所研究的数学分支的全部命题。

证明一个公理体系是完备的方法是，要证明这个公理体系的任何两个模型是同构的。这是因为如果公理体系 Σ 是不完备的，则 Σ 中允许加入与其公理独立的公理 A 使 $\Sigma + A$ 成为一个无矛盾的扩大了的公理体系，则 $\Sigma + A$ 必有一个模型 M，则 M 也是 Σ 的模型。另一方面，因为 A 有独立性，则 $\Sigma + \overline{A}$ 也是无矛盾的系统，设 M' 是 $\Sigma + \overline{A}$ 的模型，则 M' 也是 Σ 的模型。这样公理体系 Σ 有两种不同的模型 M 与 M'，它们显然不是同构的。

例如，由于欧氏几何模型和罗氏几何模型都包含了绝对几何公理体系 Σ_0，但这两模型是不同构的，故绝对几何公理体系是不完备的。

在公理体系中的三个基本问题中，相容性对任何体

系都是必须的，而独立性、完备性都是不必须的。数学中有许多重要的公理体系（例如，代数中群的公理），正是因为不具备完备性，才有各种不同构的模型，从而显示出广泛的应用。

4. 公理化方法的应用

认识对偶原理，在公理系统中是最清楚不过的了，这也是公理化方法的一种重要应用。例如，前面所述的布尔代数系统，其满足的 $1^\circ \sim 4^\circ$ 的性质，其实就是布尔代数的公理。由于每条公理都包含两部分，它们相互对偶，亦即布尔代数的公理是相互对偶的，故由布尔代数公理推出的一切结论亦必相互对偶，从而在布尔代数中，对偶原理成立。

作为应用的另外一例，我们来考察平面射影几何对偶原理的正确性。平面射影几何公理共三组，它们顺次是结合公理、顺序公理、连续公理。这里仅以结合公理为例说明。

平面射影几何结合公理：

I_1 通过两点有且仅有一条直线；

I_2 两条直线通过且仅通过一点；

I_3 存在四点，其中无三点共线；

I_4 若 $\triangle ABC$ 与 $\triangle A'B'C'$ 中对应顶点的连线 AA'、BB'、CC' 共点，则对应边的交点

$$BC \times B'C' = P, CA \times C'A' = Q, AB \times A'B' = R$$

共线（图 2.10）。

证明： 1° 显然 I_1 与 I_2 是对偶的。

2° 现证 I_3 的对偶成立。

由公理 I_3，存在无三点共线的四点 A_1、A_2、A_3、A_4，按公理 I_1 和 I_2，存在直线

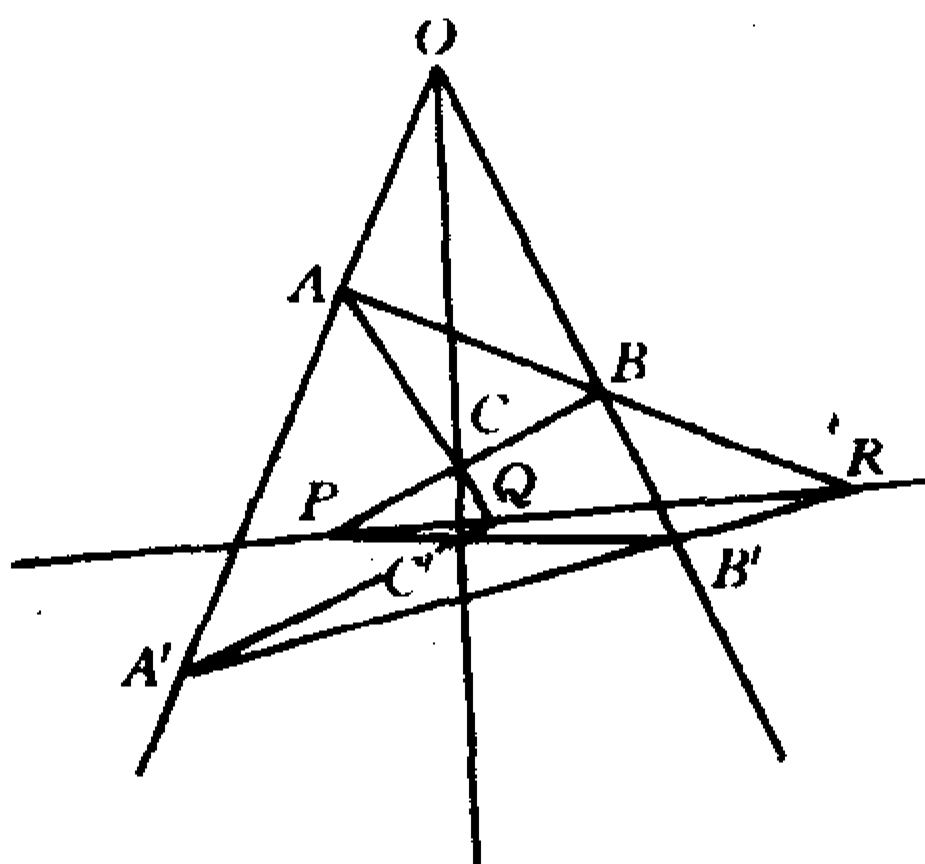


图 2.10

$l_1: A_2A_3B_1$，　$l_2: A_3A_1B_2$，　$l_3: A_1A_2B_3$，

$l_4: A_1A_4B_1$，　$l_5: A_2A_4B_2$，　$l_6: A_3A_4B_3$。

其中

$$A_2A_3 \times A_1A_4 = B_1, A_2A_4 \times A_1A_3 = B_2,$$

$$A_1A_2 \times A_3A_4 = B_3。$$

这六条直线是互异的，从而点 B_1、B_2、B_3 与 A_1、A_2、A_3、A_4 不同，而且 B_i 与 B_j 不同。例如说，若 $B_1 \equiv A_1$，则 A_1、A_2、A_3 将共线 l_1；若 $B_1 \equiv B_2$，则 A_1、A_2、A_4 共线。这样，如下的四条直线 l_1、l_3、l_4、l_6 便是无三线共点的四直线，故 I_3 的对偶成立。

3° 再证 I_4 的对偶成立，亦即它的逆命题成立。即已知 $\triangle ABC$ 与 $\triangle A'B'C'$ 中，若 $BC \times B'C' = P, CA \times C'A' = Q, AB \times A'B' = R$ 共线，则 AA'、BB'、CC' 共点（图 2.10）。

设 $AA' \times CC' = O$。考察三角形 RAA'、PCC'，因为 P、Q、R 共线，亦即 AC、$A'C'$、RP 共点于 Q。由 I_4 知：

$RA \times PC = B$、$RA' \times PC' = B'$、$AA' \times CC' = O$ 共线，亦即 BB' 过 AA'、CC' 的交点，从而 AA'、BB'、CC' 共点于 O，故 I_4 对偶成立。这样，结合公理的对偶均为真。

同样可以证明顺序公理、连续公理的对偶亦成立，故平面射影几何中对偶原理成立。

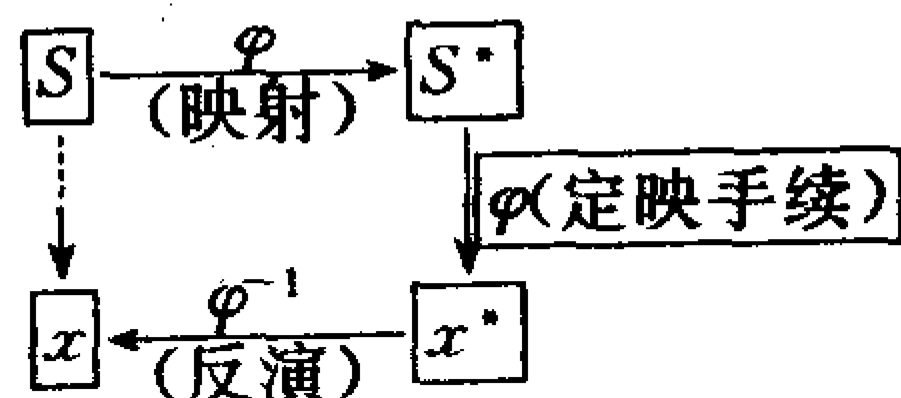
应该指出，平面射影几何公理体系中的 I_4 是笛沙格命题。熟悉射影几何的读者都知道，由于平面笛沙格命题的证明必须借助空间作图，故对空间射影几何而言，笛沙格命题可以作为定理，而对平面射影几何而言，笛沙格命题必须取作为公理，即 I_4。

2.6 关系、映射、反演思维方式

关系、映射、反演的思维方式，是处理问题的一种普遍思想原则，其实质是矛盾转移法，即将较困难的问题转化为较易处理的问题，以求得问题的解决。在数学中，它有广泛的应用。

1. 关系、映射、反演方法的含义

关系、映射、反演方法是指如下处理问题的一种思想原则：即给定一个含有目标原象 x 的关系结构 S，如果能找到一个可定映映射 φ，将 S 映入或映满 S^*，则可从 S^* 通过一定的数学方法(定映手续)把目标映象 $X^*=\varphi(x)$ 确定出来，进而通过反演 φ^{-1} 又可把 $x=\varphi^{-1}(x^*)$ 确定出来，这样，原来的问题就得到了解决。用框图表示是：



其过程有如下的步骤：

关系 → 映射 → 定映 → 反演 → 得解

由于关系、映射、反演的英文分别为 Relationship、

Mapping、Inversion，故这一方法又常简记为 RMI 方法。

以上叙述中，所谓关系结构是指彼此具有某种或某些数学关系的数学对象的集合。其中，数学对象则是泛指各个具体数学理论中所涉及的数学概念。如数、量、向量、函数、方程、泛函、点、线、面、几何图形、空间、集合、运算、算子、群、环、域、范畴等，其共同特征是具有量性、逻辑构造性和一义确定性。而数学关系即是数学对象之间的确定关系，如代数关系、序关系、拓扑关系、函数关系等。凡在两类数学对象或两个数学集合的元素之间建立了一种“对应关系”，则就说定义了一个映射。显然这里所说的“映射”一词，与集合论中的“映射”概念相比，已赋予了更为广泛的含义。若映射 φ 是可逆的，则其逆映射 φ^{-1} 称为“反演”。

例如，求椭圆 $\frac{x^2}{a^2}+\frac{y^2}{b^2}=1$ 的面积，作仿射变换

$$\varphi:\begin{cases} x'=\frac{x}{a},\\ y'=\frac{y}{b},\end{cases}$$

即

$$\begin{pmatrix} x'\\ y'\end{pmatrix}=\begin{pmatrix}\frac{1}{a} & 0\\ 0 & \frac{1}{b}\end{pmatrix}\begin{pmatrix} x\\ y\end{pmatrix}。$$

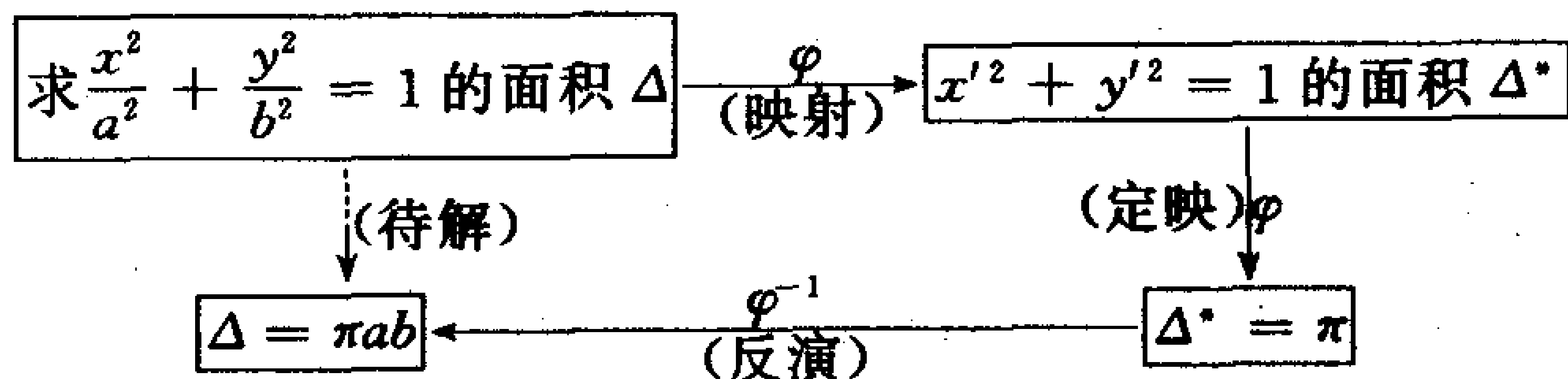
则椭圆变为圆 $x'^2+y'^2=1$。若分别以 Δ、Δ^* 表示椭圆、圆及它们的面积，这里椭圆方程及其面积等数量关系的全体便是关系结构 S，Δ 是目标原象；圆方程、面积等的全体便是映象关系结构 S^*，Δ^* 是目标映象。又因

$$\begin{vmatrix} \frac{1}{a} & 0 \\ 0 & \frac{1}{b} \end{vmatrix} = \frac{1}{ab} \neq 0$$

是仿射变换的变积系数，故

$$\Delta = \Delta^* \Big/ \frac{1}{ab} = \pi ab。$$

这里，确定目标映象的数学方法，亦即定映手续 φ 便是圆面积公式，其求解的过程为：



亦即

$$(S,\Delta) \xrightarrow{\varphi} (S^*,\Delta^*) \xrightarrow{\psi} \Delta^* \xrightarrow{\varphi^{-1}} \Delta。$$

2. 关系、映射、反演方法的渊源及其使用要点

如同数学定理的发现和数学理论的建立一样，数学思维方式的产生也是数学家们在数学研究的实践中不断总结、提炼而形成的。关系、映射、反演方法，就是在科学方法论的“化归”原则的基础上，经进一步抽象、提炼、加工后产生的。

历代的数学家，在数学研究的实践中一直十分注意“化归”原则的应用，亦即是注意把问题进行由未知到已知、由难到易、由复杂到简单的转化。“数学之神”阿基米德就是最杰出的代表之一。如同马塞罗斯的传记作者洛塔奇所说：“阿基米德处理的问题是最复杂而又最困难的，然而他用于解决这些问题的方法却是最简单而又最明了的。”阿基米德晚年在一篇题为《方法》的短文中，总

结了他数学研究的方法指出:“力学便于我们发现结论,而几何则能帮助我们对结论作出证明。”他坚信:“在现在或未来的研究者中,一定会有人利用我这里提出的方法获得我还不曾想到的其他的定理。”[21] 阿基米德自己曾利用力学的平衡原理解决了诸如球体积公式、抛物线弓形面积等问题,其思想方法就是化归。可以同时称为哲学家和数学家的笛卡儿,1637 年发表了他的方法论专著,提出了一种自称为“万能”的方法:

第一,把任何问题化归为数学问题;

第二,把任何数学问题化归为代数问题;

第三,把任何代数问题化归为方程式求解。

显然,笛卡儿的所谓万能模式不是万能的,但有其重要应用,这也正是他赖以建立解析几何的基本思想原则。笛卡儿以后,英国哲学家霍布斯对此论述道:“从一个愿望联想起我们曾经看到过的某些方法与手段,借助这些方法与手段,我们可以得到如所求之目标那样一类东西。”他还提出了“推理我意谓计算”的思想,他的这两段论述,实际上可看作化归原则的具体阐述。这种思想对后来数理逻辑的产生和发展具有重要意义。对化归原则进行严格数学刻画的是关系、映射、反演方法,它是通过数学对象、关系结构、映射、定映、反演、目标原象、目标映象、原象结构、映象结构等科学语言,并借助直观、明晰的框图结构来完成的,这种刻画标志着化归原则达到了新的更高的抽象程度和更好的严密程度,从而也就具有更广泛的应用,故从化归原则到 RMI 原则的发展,是数学方法的一次重要进步。在这一进步中作出贡献的是我国数学家徐利治,是他在 1983 年初版的《数学方法论选讲》中首先提出了这一原则。

在 RMI 思维方式的五个组成部分中，除包含目标原象 x 的关系结构是原先给定的之外，其余四个部分都由选取的 φ 而定，因此 φ 的选取成为 RMI 思维方式的核心。φ 的选取要求是要使得定映手续 ψ 存在，且能容易地通过 ψ 定出 x^*；这样的选择是不容易的，因而它是使用 RMI 方法的关键和困难之所在。如果针对一大类重要问题，真能找到一个统一的、普遍的映射法，那对数学将是一种重要的贡献。

3. 关系、映射、反演方法的应用

由于 RMI 方法是一种处理问题的普遍思想方法，因而具有广泛的应用，它不仅可作为探求问题证明的一种重要思路，而且还可以作为数学创造的一种方法原则。

例 1 RMI 方法应用的广泛性。

(1) 对数方法。

一个含有乘、除、乘方、开方的复杂计算式，通过对数化为加、减运算。这里，映射 φ 是取对数，定映是查对数表计算，反演是取反对数，故对数方法是 RMI 方法的具体应用。其意义十分重大，拉普拉斯曾称道说："对数计算通过缩短计算的时间，而延长了天文学家的生命。"伽里略还甚至说："给我空间、时间及对数，我即可创造一个宇宙。"

(2) 解析几何方法。

解析几何中含目标原象的关系结构是几何问题，映射 φ 就是建立坐标系进行解析表示，映象关系结构便是相应的代数问题。从代数结论中看出几何解释便是反演 φ^{-1}。故解析几何方法亦是 RMI 方法的具体表现。对解析几何的发明，恩格斯评价道："数学中的转折点是笛卡儿的变数。有了变数，运动进入了数学，有了变数，辩证法进

入了数学，有了变数，微分和积分也就立刻成为必要的了，而它们也就立刻产生……"[6]

(3) 数学模型方法。

数学模型方法中映射 φ 是从现实原型到数学模型的某种对应关系，模型中得出的数学结论回到现实原型上就是反演，故数学模型方法实质亦是 RMI 方法的表现。

例 2 代数问题映射为几何问题。

任给 8 个非零实数 $a_1,a_2,\cdots,a_8$。证明：六个数 $a_1a_3+a_2a_4,a_1a_5+a_2a_6,a_1a_7+a_2a_8,a_3a_5+a_4a_6,a_3a_7+a_4a_8,a_5a_7+a_6a_8$ 中至少有一个是非负的。

注意到

$$\begin{aligned}&2(a_1a_3+a_2a_4)\\&=|a_1+a_2i|^2+|a_3+a_4i|^2-|(a_1-a_3)\\&\quad+(a_2-a_4)i|^2\\&=|a_1+a_2i|^2+|a_3+a_4i|^2-|(a_1+a_2i)\\&\quad-(a_3+a_4i)|^2,\end{aligned}$$

故作映射 φ：

$$a_1、a_2\to\overrightarrow{OA}:a_1+a_2i;$$
$$a_3、a_4\to\overrightarrow{OB}:a_3+a_4i;$$
$$a_5、a_6\to\overrightarrow{OC}:a_5+a_6i;$$
$$a_7、a_8\to\overrightarrow{OD}:a_7+a_8i。$$

这样

$$a_1a_3+a_2a_4\geqslant 0\Leftrightarrow|\overrightarrow{OA}|^2+|\overrightarrow{OB}|^2-|\overrightarrow{BA}|^2\geqslant 0。$$

由于四个向量两两交角中，至少有一个不超过 $\frac{360°}{4}=90°$，不妨设 $\overrightarrow{OA}$、$\overrightarrow{OB}$ 的夹角不超过 90°，在 $\triangle AOB$ 中应用余弦定理

$$|\overrightarrow{OA}|^2+|\overrightarrow{OB}|^2-|\overrightarrow{AB}|^2\geqslant 0,$$

从而

$$a_1a_3 + a_2a_4 \geqslant 0。$$

故六个数 $a_1a_3 + a_2a_4, a_1a_5 + a_2a_6, a_1a_7 + a_2a_8, a_3a_5 + a_4a_6, a_3a_7 + a_4a_8, a_5a_7 + a_6a_8$ 中至少有一个是非负的。

这里，目标原象是一个代数问题，通过映射 φ，化为一个几何问题，定映手续 ψ 是余弦定理，证题的思想体现了 RMI 方法。

例 3 三阶幻方的个数。

三阶幻方，用 1,2,…,9 九个数字组成，幻和为 15。若 9 个位置按行的顺序依次为 $x_1, x_2, \cdots, x_9$，则按行、列两对角线及幻和的顺序排列，有方程组

$$AX = B,$$

其中

$$X = (x_1, x_2, \cdots, x_9)^T,$$

$$B = (15, 15, \cdots, 15, 45)^T,$$

$$A = \begin{bmatrix} 1&1&1&0&0&0&0&0&0 \\ 0&0&0&1&1&1&0&0&0 \\ 0&0&0&0&0&0&1&1&1 \\ 1&0&0&1&0&0&1&0&0 \\ 0&1&0&0&1&0&0&1&0 \\ 0&0&1&0&0&1&0&0&1 \\ 1&0&0&0&1&0&0&0&1 \\ 0&0&1&0&1&0&1&0&0 \\ 1&1&1&1&1&1&1&1&1 \end{bmatrix}。$$

将 (A, B) 进行初等变换，使对角线元素化为 1，而得

$$\left[\begin{array}{ccccccccc:c} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 10 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 10 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & -1 & -1 & -5 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & -2 & -10 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 5 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 2 & 20 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 15 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

则 $AX = B$ 的解为

$$\begin{cases} x_1 = 10 - x_9 \\ x_2 = 10 - x_8 \\ x_3 = -5 + x_8 + x_9 \\ x_4 = -10 + x_8 + 2x_9 \\ x_5 = 5 \\ x_6 = 20 - x_8 - 2x_9 \\ x_7 = 15 - x_8 - x_9。 \end{cases}$$

上式中 x_8、x_9 可以是任意的，但由于要求 $(x_1, \cdots, x_9)$ 是 $(1,2,\cdots,9)$ 的一个排列，故只有当 (x_8, x_9) 为 $(1,6)$；$(1,8)$；$(3,4)$；$(3,8)$；$(7,2)$；$(7,6)$；$(9,2)$；$(9,4)$ 时有解。但这 8 个解都和图 2.4 中的三阶幻方同构。如 (x_8, x_9) 为 $(1,6)$ 时，得图 2.4 中的幻方，而当 (x_8, x_9) 为 $(1,8)$ 时，所对应的幻方可由图 2.4 中的幻方按中间列反射而得。像这样一个可由另一个经反射或旋转后得到的两个幻方，称为是同构的。故三阶幻方在同构的意义下只有一个。

上述解法中的映射是将幻方问题对应一个 9×10 矩阵后求解的，定映的手续是矩阵的初等变换。

2.7 反例思维方式

举反例也是数学的重要思维方式，可以说反例与证明同样重要，它是一个问题的两个侧面。如同数学家B.R. 盖尔鲍姆，J.M.H 奥姆斯特德所指出的："数学由两大类 —— 证明和反例组成。而数学发现也是朝着两个主要的目标 —— 提出证明和构造反例。"[22]

1. 反例的实质及其作用

数学中的反例通常是指推翻某个命题成立的例子。例如，若 P、Q 为简单命题，由真值表知，复合命题"若 P 则 Q"假，当且仅当 P 真且 Q 假。因而推翻复合命题"若 P 则 Q"成立的反例就是 P 真且 Q 假中的一个例子。显然，反例的实质就是证伪。即要证明"若 P 则 Q"为假，当举出了 P 真且 Q 假的某一例子后，便达到了证伪的目的。

反例根据数学命题的构成可分为不同的类型，主要的有如下几种。

(1) 简单命题的反例。

性质型简单命题，有四种基本形式，即全称肯定判断，全称否定判断，特称肯定判断和特称否定判断，则断定全称肯定判断(所有 S 都是 P)为假的反例是特称否定判断(有 S 不是 P)，而断定全称否定判断(所有 S 都不是 P)为假的反例，是特称肯定判断(有 S 是 P)。用量词表示为：

$\forall xF(x)$，其反例形式为 $\exists x\overline{F(x)}$；(1)

$\forall x\overline{F(x)}$，其反例形式为 $\exists xF(x)$。(2)

例如，1640年，费玛(Fermat)提出了"形如 $2^{2^n}+1$(n 为非负整数)的数都是素数"的猜想。1732年，欧拉指出：

当 $n=5$ 时，

$$2^{2^5}+1=4294967297=641\times6700417$$

是合数。欧拉对 $n=5$ 的例子就是推翻费玛猜想的反例。前者是全称肯定判断，后者是特称否定判断。

(2) 充分条件、必要条件的反例。

要断定"P 是 Q 的充分条件"为假的反例，是 P 真且 Q 假的例子，亦即 $P\wedge\overline{Q}$ 的例子。断定"P 是 Q 的必要条件"为假的反例，是 Q 真且 P 假的例子。因为 P 是 Q 的必要条件，则 Q 是 P 的充分条件，从而反例是 Q 真且 P 假的例子。例如，要断定"函数在一点可微是其在该点连续的必要条件"为假，只要找出函数在一点连续但不可微的例子。比如，反例 $y=\sqrt[3]{x}$，它在 $x=0$ 处连续但不可微。

(3) 条件变化的反例。

许多数学命题，其条件有多个，这时条件变化(增加、减少或更换)时导致命题为假的反例就是条件变化的反例。

例如，数学分析中，罗尔定理中有条件：

1° $f(x)$ 在 $[a,b]$ 上连续；

2° $f(x)$ 在 (a,b) 内可导；

3° $f(a)=f(b)$。

结论是：至少存在 $\xi\in(a,b)$，使 $f'(\xi)=0$。这三个条件缺一不可。针对不满足某一条件定理为假，都可制作不同类型的反例。如对条件 1° 不满足时有反例：$y=x-[x]$，$x\in[0,1]$。这例中，条件 2°、3° 成立，但罗尔定理不成立。因为它在除 $x=1$ 以外连续。又对条件 3° 不满足时有反例：$y=x$，$x\in[0,1]$。这时条件 1°、2° 满足，但

$f(0) \neq f(1)$，罗尔定理不成立。

反例的作用是十分明显的：

(1) 通过反例，可以发现数学中原有理论的局限性和不足，从而推动数学不断向前发展。

数学史上这样的例子是不胜枚举的。例如，公元前 5 世纪，毕达哥拉斯学派的希帕索斯(Hippasus) 发现了等腰直角三角形的直角边与斜边无公度，其实质是该学派关于宇宙间的一切数量关系都可以归结为整数或整数比的反例。这一反例的深刻意义，如同 M. 克莱因所指出的，“不可公度长的发现对 Pythagoras 派哲学是那么可悲的打击：不可公度长之比竟然不能用整数之比来表示。此外，他们相信一直线是由有限个点(他们把它和物理质点视为等同) 组成的；但对 $\sqrt{2}$ 那样的长就不可能是如此。如果他们把无理数当作数来接受，他们那个视整数为至上的哲学就要垮台。”[17] 在反例的推动下，无理数进入了人们的研究领域。公元前 3 世纪，欧几里得在《几何原本》的第十篇中讨论了几何代数法里出现的无理量，又在第十篇的第三章中给出了 $\sqrt{2}$ 为无理数的证明。19 世纪后期，数学家们已不再把连续性与可微性混为一谈，但他们仍相信，连续函数除个别点外总是处处可微的。1872 年，当时已是 57 岁的外尔斯特拉斯构作了如下处处连续却处处不可微的函数

$$f(x) = \sum_{n=1}^{\infty} b^n \cos(a^n \pi x), \quad x \in (-\infty, +\infty)。$$

其中，a 是奇整数，b 是 0 与 1 间的实数，而 $ab > 1 + \frac{3}{2}\pi$。其图象是一条连续不断的曲线，但处处见棱见角。外尔斯特拉斯的反例，促使人们认真检查分析的理论基础，从而

形成了尔后的“分析算术化”运动，对数学的继续发展有着深远的影响。

又如，1779年，欧拉发表了一篇题为《有关一类新幻方》的论文，文中他引进了被后人称做为正交拉丁方的概念，用以构作幻方。所谓拉丁方是一种 $n \times n$ 方阵，若用 n 种符号填进这方阵后，每行每列都出现全部符号。两个 n 阶拉丁方若对应位置的字母组成数偶，这样 n^2 个数偶均不同，就称为正交的。对正交拉丁方，欧拉断言：“我毫不迟疑地下这样的结论，不可能构作一对正交的6阶拉丁方，而且这结论亦可推广至10阶、14阶……的情形，即是当阶数是二乘一个奇数的情形。”[23] 本世纪50年代，原籍印度的美国数学家玻色(Bose)和他的学生西里克汉特(Shrikhande)找到了一对22阶的正交拉丁方，随后美国数学家派克(Parker)亦找到了一对10阶正交拉丁方，这些反例推翻了欧拉的猜想。不久，他们还合作证明了一个极为漂亮的定理：除 $n=2$ 和6以外，总存在一对正交的 n 阶拉丁方。正交拉丁方的研究，推动了 n 阶射影平面存在性的讨论。1938年玻色证明了一个 $n(n>2)$ 阶射影平面相当于 $n-1$ 个两两正交的 n 阶拉丁方。至目前，$n=10,12,14,15,18,\cdots$ 阶射影平面是否存在尚属未知。例如，$n=10$ 的问题，被归结为是否有9个两两正交的10阶拉丁方问题。迄今连3个两两正交的10阶拉丁方也还没有找到。可见，反例对人们思维的刺激是何等的强烈。

(2) 反例对理解和深化概念具有重要意义。

一个正确的认识往往要经过正反两方面的比较和鉴别才能确立。反例通过证伪从反方向帮助人们理解概念和加深对概念的认识。例如，外尔斯特拉斯的著名反例，

对深化人们搞清连续与可微间的关系是不可多得的。罗尔定理的条件型反例，对加深人们对定理的理解也是很有帮助的。又如，关于周期函数的最小正周期，人们常认为，一个周期函数必有最小正周期，但如下的反例却澄清了这一认识：

$$设\ f(x)=\begin{cases}1, x\ 为有理数,\\ -1, x\ 为无理数。\end{cases}$$

这是以任何有理数T为周期的函数。因为x为有理数时，$x+T$也为有理数，x为无理数时，$x+T$亦为无理数，从而

$$f(x+T)=\begin{cases}1, x\ 为有理数。\\ -1, x\ 为无理数。\end{cases}$$

但有理数T无最小正有理数，故$f(x)$是周期函数但无最小正周期。

2. 反例的构造方法

构造反例是一种从无到有的创造，它对人们思维素质的锤炼和数学能力的培养都具有重要的意义。

反例的构造可从如下一些方面入手进行。

(1) 通过寻找反例成立的范围来构造反例。

若P、Q为两个简单命题，其外延分别为集合A、B，则“若P则Q”的证明转化为$A\subseteq B$的证明。而证明命题不成立的反例就是要构造$A\cap\overline{B}$中的例子，故$A\cap\overline{B}$就是反例成立的范围。当寻找到了这一范围，反例也就得到了构造。

例 1 试构作命题“过圆锥的两条母线所作的一切截面中，以轴截面的面积最大”的反例。

设PAB为过圆锥的两条母线的截面，顶角为β，PBC是轴截面，顶角为α(图 2.11)。

要构造结论不成立的反例，即要求

$$S_{\triangle BPC}=\frac{1}{2}l^2\sin\alpha$$
$$<\frac{1}{2}l^2\sin\beta=S_{\triangle BPA}$$
$$\Leftrightarrow\sin\alpha<\sin\beta。$$

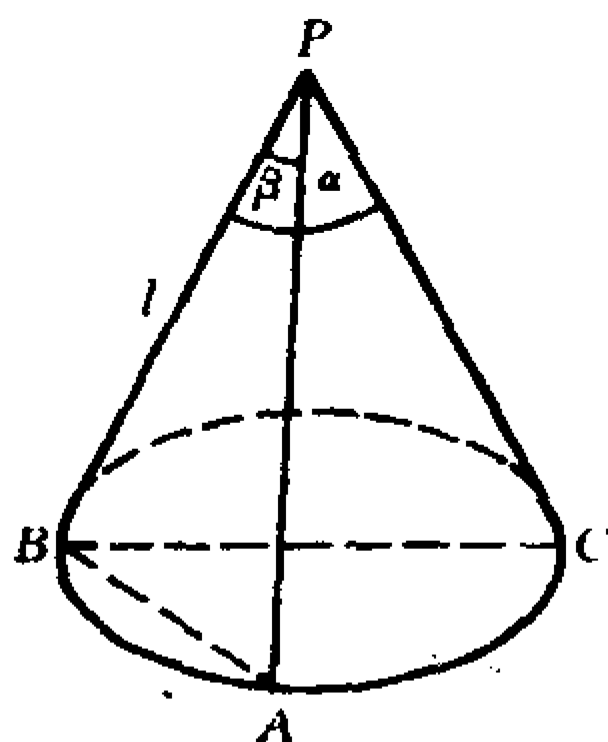


图 2.11

所以，$0<\alpha<\beta\leqslant\frac{\pi}{2}$，或 $\frac{\pi}{2}\leqslant\beta<\alpha<\pi$。

若取 $\alpha=\frac{2}{3}\pi,\beta=\frac{\pi}{2}$ 便得反例：

$$S_{\triangle BPC}=\frac{\sqrt{3}}{4}l^2<\frac{1}{2}l^2=S_{\triangle BPA}。$$

从而，给出的命题“过圆锥的两条母线所作的一切截面中，以轴截面的面积最大”是不成立的。

例 2　已知命题“设一个三角形的三个角和两边与另一个三角形的三个角和两边分别相等，则这两个三角形全等”。试构造其证伪的反例。

设一三角形的三边为 $a<b<c$，另一三角形的三边为 $b<c<d$。由条件知两三角形是相似的。为构造反例，设

$$\frac{b}{a}=\frac{c}{b}=\frac{d}{c}=q(q>1)。$$

注意到 $a+b>c$，即 $a+aq>aq^2$，或 $q^2-q-1<0$。

则有　$$1<q<\frac{1+\sqrt{5}}{2}。$$

故反例成立的范围是 $1<q<\frac{1+\sqrt{5}}{2}$。

取 $q=\frac{n+1}{n}(n>2)$，$a=n^3$ 便得反例：

$$a = n^3, b = n^2(n+1), c = n(n+1)^2;$$
$$b = n^2(n+1), c = n(n+1)^2, d = (n+1)^3。$$

这两三角形三角、两边分别相等，但图形不全等。

本例说明三角形全等判定中，“对应”两字是不可忽视的。

(2) 从考察问题的极端情况中构造反例。

具有一定性态的数学对象，常在某些极端情况发生质变。故构造“若 P 则 Q”的反例，常考虑命题 P 所断言的对象在边界点、奇点、极限点等的特殊情况下，命题 Q 所断言的性态的变化入手。

例如，$y = |x|$ 处处连续，但在 $x = 0$ 处不可微。对这一反例的构造，正是着眼于函数 $y = |x|$ 在 $x = 0$ 这一奇异点上的性态变化。

(3) 类比已知反例构造新反例。

数学中相伴出现的一些对象，如低次与高次，一元与多元，低维与高维等，它们既有区别，又有联系，这种情况为类比方法的应用提供了广阔的背景，从而也为反例的构造提供了方法。

设 $x_1 \in A_1 \cap \overline{B_1}$ 为“若 P_1 则 Q_1”的反例（A_1、B_1 为 P_1、Q_1 外延的集合）。又“若 P_1 则 Q_1”的类比物是“若 P_2 则 Q_2”，x_2 又是 x_1 的类比物，则 $x_2 \in A_2 \cap \overline{B_2}$ 往往成立（A_2、B_2 是 P_2、Q_2 外延的集合）。亦即构作“若 P_2 则 Q_2”的反例可通过类比“若 P_1 则 Q_1”的反例来进行。

例 3 “空间一点 P 到一个凸四面体四个顶点的距离相等，则 P 点的位置必定在形内”。试构造其证伪的反例。

凸四面体的类比物是平面凸四边形。故本题在平面内的类比命题是：

"平面一点 P 到一个凸四边形四个顶点的距离相等，则 P 点的位置必定在形内。"

已知平面内的上述命题证其伪的反例是：

以线段 AB 为直径作半圆，在半圆周上取两个异于 A、B 的点，并与 A、B 相连得一四边形，则 AB 之中点 P 到四边形各顶点的距离相等。这时，P 在四边形的边上。若在上述半圆周上取不同于 A、B 的相异四点得一四边形，则 P 在四边形外。若所取的四点是以 AB 为直径的整个圆的内接正方形的四个顶点，这时 P 在四边形内。亦即在前两种情况下，是否定上述命题的反例。

用类比法构作本例问题的反例为：

设有一球，在球的任一过球心的截面大圆上任取三点 A、B、C，并使这三点三等分大圆，再在半球面上任取一点，则球心至四点的距离相等，而球心 P 在四面体的面 ABC 上。若所取的四点，均在过球心大圆截面的一侧，这时球心 P 至四点等距，但 P 在四面体外部。若所取的四点是球的内接正四面体的四个顶点，此时球心 P 在四面体的内部。而前两种情况，就是否定本例命题的反例。

(4) 借助直观构造反例。

若要构造的反例所涉及的内容有比较明显的几何特征时，这时构造工作可借助直观来进行。

例 4　"已知 T_1、T_2 分别是 $f(x)$、$g(x)$ 的最小正周期，则 T_1、T_2 的最小公倍数是 $f(x)+g(x)$ 的最小正周期"。试构造否定这命题的反例。

由于函数的周期性有明显的几何特征，故可借助直观进行构造。

我们设计当 $T_1=T_2=T$ 时，$f(x)+g(x)$ 的周期反而缩小了。如图 2.12(1)、(2) 中 $f(x)$、$g(x)$ 周期均为 T，

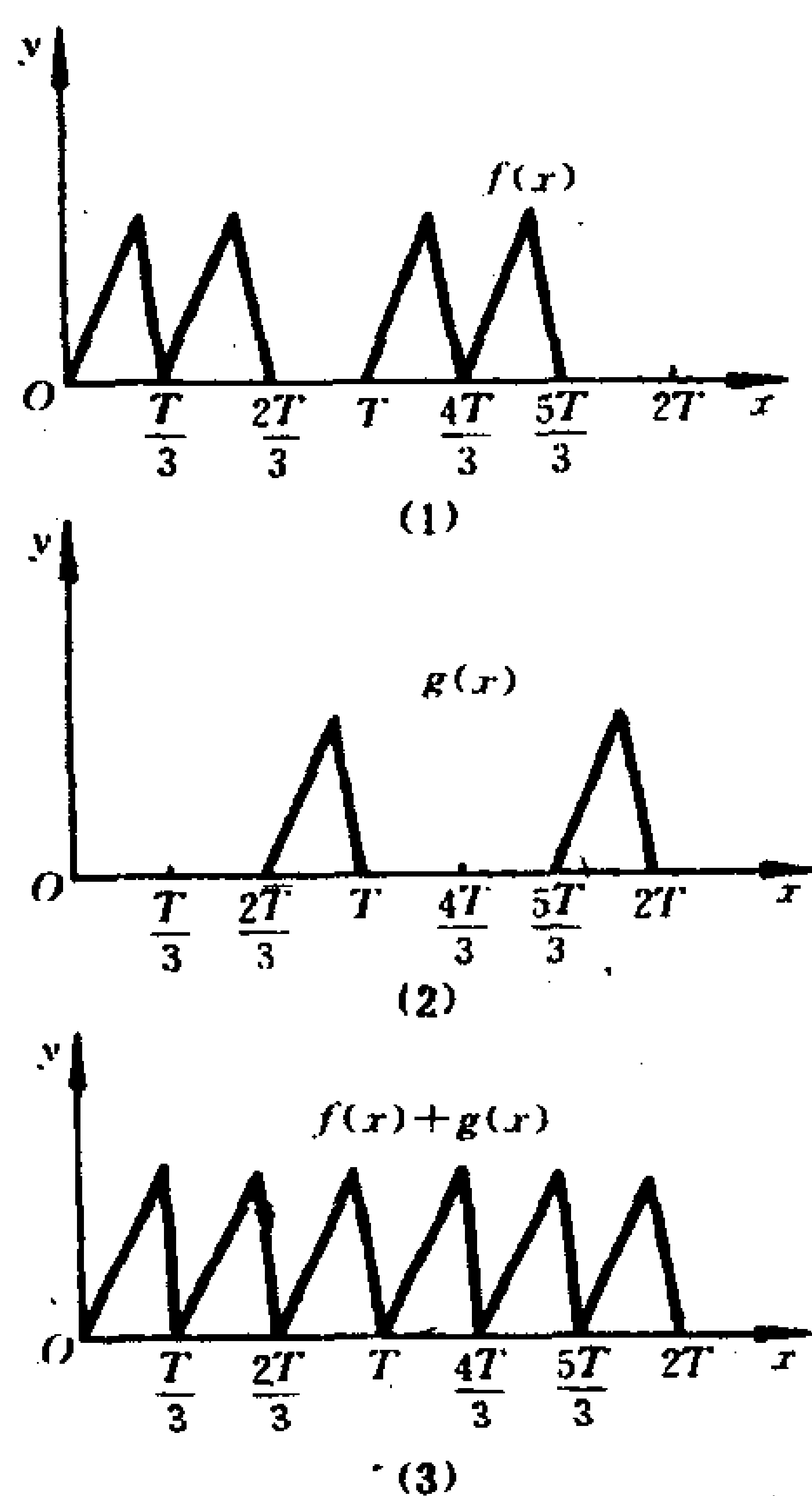


图 2.12

但 $f(x)+g(x)$ 周期却为$\frac{T}{3}$(图 2.12(3))。$f(x)$ 在 T 内的图象有三段组成,前两段是相同的一升一降的两段折线,第三段重合于 x 轴。而 $g(x)$ 前两段重合于 x 轴,第三段是一升一降的折线。

三　数学中的具体思维原理、原则、方法

思维方法是思维方式的重要组成部分，是主体思维活动为实现一定思维目的所采用的规则、手段、途径和技能技巧构成的综合体。数学中的概念、原理、原则和方法，蕴含着极为丰富、深刻的思想，它是数学思维方法因而也是数学思维方式的重要组成部分。研究数学中的思维方式和方法，离不开对数学的重要概念、原理、原则和方法的探讨。本章我们将展示在数学、自然科学、社会科学中起基础作用的若干数学概念、原理、原则和方法，揭示其深刻的思想内容，探讨其在“外部世界的科学”和数学“自己的研究领域”中的重要应用。

3.1　集合论思想与计数原理

1. 康托的集合论思想

集合论是由德国数学家乔治·康托(Georg Cantor)创立的，其基本原则或思想方法，不外乎概括原则、外延原则、一一对应原则、对角线方法及其实无穷思想。

(1) 概括原则与实无穷思想。

概括原则是指：对于任一性质 p，我们就能且仅能把所有满足给定性质 p 的对象 g 汇集在一起而构成一个集合 G 的这样一种造集原则。用符号表示即为：

$G=\{g|p(g)\}$ 或 $\forall g(g\in G\leftrightarrow p(g))$。

关于“集合”本身的含义，康托曾作过描述说：“把一些明确的(确定的)、彼此有区别的、具体的或想象中抽象的东西看作一个整体，便叫作集合。”[24] 这里，康托用整体来说明集合，不能称为集合的定义，实际是一种同义反复。迄今为止，一切想要对集合作出所谓严谨的、合乎数学要求的定义的尝试都没有成功，以致近代公理集合论者，都放弃了对集合下定义的做法，而把它作为基本概念用公理加以隐定义。

概括原则的采用，立即在数学中引入了实无穷思想，因为满足条件 $p(g)$ 的 g 当然可以是无限多的。

(2) 外延原则。

外延原则是指两个集合 A、B，当且仅当它们的元素完全相同时，才把它们看作是相同的，即

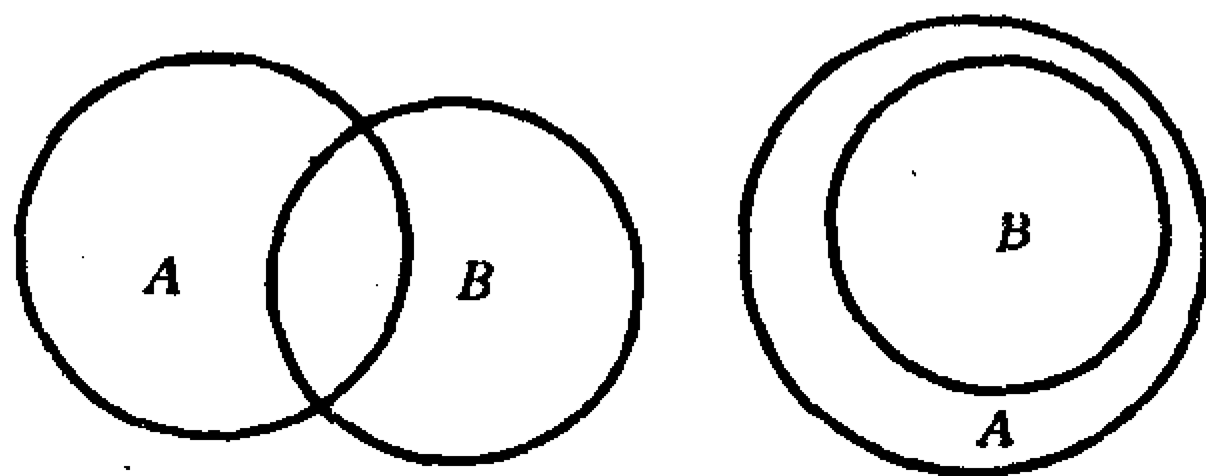
$$A = B \Leftrightarrow \forall\ x(x \in A \leftrightarrow x \in B)。$$

根据外延原则可产生集合的包含关系和并、交、差等运算，其性质类似于数的加、乘、减运算，但也有很多不同之处。例如，

$$(A' - B) \cup B = A \cup B。$$

只有在 $B \subseteq A$ 的特殊情况之下才有

$$(A - B) \cup B = A(见图 3.1)。$$



$(A - B) \cup B = A \cup B$　$(A - B) \cup B = A$

图 3.1

如果说康托的概括原则保证了各种不同集合的存在

性，那么外延原则则保证了集合的确定性。

(3) 一一对应原则与对角线方法。

一一对应原则是指，若集合 A 与集合 B 的元素之间建立了一一对应关系，则集合 A 与 B 有相同的势(或基数)，记为$\bar{\bar{A}} = \bar{\bar{B}}$ 或 $|A| = |B|$，并称这样的两个集合是对等的，记作 $A \simeq B$。这是康托 1878 年在《克莱尔杂志》上发表的题为《集合论》的文章中提出的。这一原则揭示了无穷集的这样一个本质特征，即其整体与部分在数量上是可以处于同等地位的。例如，自然数集与其平方数集在元素的个数上对等，但后者却是前者的部分。所说的一一对应原则和对角线方法结合起来还能进一步揭示如下事实：即无穷集合不是清一色的，它既存在像自然数集N那样的可数无穷集，又有实数集那样的不可数无穷集。其中前者的势记为 $\aleph_0$(阿列夫零)，后者的势记为C，称为连续统势。而有理数集也是可数无穷集，$[0,1]$ 中的全体实数是不可数无穷集，这些都可用康托的对角线方法证明。对于不同无穷集基数大小的比较，康托的思想是：若无穷集 A 与 B，当 B 与 A 的一个子集构成一一对应，而 A 却不能与 B 的任何子集构成一一对应，则称 A 的基数大于 B 的基数。他认定最小的无穷势是 $\aleph_0$，而 C 是否是紧跟着 $\aleph_0$ 后的无穷势？康托认为答案是肯定的，即不存在任何势 m 能使 $\aleph_0 < m < C$，但无法证明。这就是著名的连续统假设。

对康托集合论思想的深层次理解，不能不涉及关于罗素悖论的问题，这是罗素在 1901 年发现的。

所有集合按其是否属于它本身，可分类为“本身分子集”(“非常集”) 和“非本身分子集”(“正常集”)。前者，集合自身也是该集合的一个元素。例如，“一切概念组成的

集合”，由于它本身也是一个概念，所以也是该集合的一个元素。后者，集合自身不是它的一个元素。例如，“自然数集”，它本身不是某个自然数，故它是“非本身分子集”。

现考虑“一切非本身分子集的集合 S”，S 属于哪一类？设 S 是本身分子集，则 S 为其自身的一个元素，而由 S 的定义，它的每一元素都是非本身分子集，则 S 是非本身分子集。又设 S 是非本身分子集，而一切非本身分子集已都包含在 S 之中了，所以 S 也一定在 S 之中，按定义，S 是本身分子集。即无论怎样都有矛盾，这就是著名的罗素悖论。

现在大多数数学家认为集合论悖论的出现，原因在于利用概括原则造集的任意性太大，因而立足于修改概括原则，即对概括原则的造集的任意性加以适当的限制，而去排除悖论的出现，这就导致了近代公理集合论的发展。例如，ZFC 系统就是以经典二值逻辑为逻辑工具，并以“集合”、“属于”为基本概念，再加上外延、空集、对偶、并集、幂集、子集、无限、选择、代换、正则等 10 条非逻辑公理构成的。在这系统中，能排除至今已出现的那些悖论，但尚不能保证在这系统中永不出现新的悖论，因为 ZFC 系统本身的无矛盾性并没有被证明。

目前，集合论已经成为现代数学的理论基础。例如，近代实变函数论的发展是以集合论作为其前提的。没有集合论，显然不会有近代的测度论，不会有描述性的实变函数论。又如，在近代数学发展中占有重要地位的抽象空间理论，无非都是具有各种特殊结论的无限集。在代数中，近世代数讨论的是具有某些结合规律的元素系统的构造，不仅以集合论概念作为其基础，而且还渗透了集合论的许多思想方法。正如亚历山大洛夫所指出的：“集合

的思想与概念已经渗透到所有的数学分支，并且改变了它们的面貌，所以不熟悉集合论的原理就不可能对近代数学获得正确的理解。"[25] 这些都表明，康托创立集合论的意义是十分重大的，借用罗素的话表述：它"可能是这个时代所有能夸耀的最巨大的工作"。

例1　外延原则的应用。

设全集 $I=\{x|x$ 为小于 20 的正偶数$\}$，若 $A\cap\overline{B}=\{12,14\}$，$\overline{A}\cap B=\{2,4,16,18\}$，$\overline{A}\cap\overline{B}=\varnothing$，求集合 A、B。

考虑由外延原则建立的集合运算性质：

$$\overline{B}=\overline{B}\cap(A\cup\overline{A})=(\overline{B}\cap A)\cup(\overline{B}\cap\overline{A})$$
$$=(\overline{B}\cap A)\cup\varnothing=\overline{B}\cap A=\{12,14\},$$

故 $B=\{2,4,6,8,10,16,18\}$。

同理，$\overline{A}=\overline{A}\cap(B\cup\overline{B})=(\overline{A}\cap B)\cup(\overline{A}\cap\overline{B})=\overline{A}\cap B=\{2,4,16,18\}$，故 $A=\{6,8,10,12,14\}$。

考虑的方法还可如下入手：

$$(A\cap\overline{B})\cup(\overline{A}\cap B)\cup(A\cap B)\cup(\overline{A}\cap\overline{B})$$
$$=[(A\cap\overline{B})\cup(\overline{A}\cap B)]\cup\{[(A\cap B)\cup\overline{A}]\cap[(A\cap B)\cup\overline{B}]\}$$
$$=[(A\cap\overline{B})\cup(\overline{A}\cap B)]\cup[(B\cup\overline{A})\cap(A\cup\overline{B})]$$
$$=[(A\cap\overline{B})\cup(\overline{A}\cap B)]\cup[\overline{(\overline{B}\cap A)\cup(\overline{A}\cap B)}]$$
$$=I。$$

今 $\overline{A}\cap\overline{B}=\varnothing$，故从 I 中除去 $A\cap\overline{B}$ 和 $\overline{A}\cap B$ 的元素后，余下的即是 $A\cap B$ 的元素，则

$$A\cap B=\{6,8,10\}。$$

因 6、8、10 这三个元素既在 A 中又在 B 中，将它们写入 $A\cap\overline{B}$ 中，便得 A 的元素。将它们写入 $\overline{A}\cap B$ 中，便得 B 的元素。

例 2 对角线方法思考之案例。

证明实数集的不可数性，康托的方法是首先证明[0，1]上的全体实数集 M 是不可数集，否则，若设 M 为一可数集，则 M 应与自然数集 N 对等，即 $M \simeq N$。于是它们间可建立一一对应关系。若[0,1]中的每一实数表成无限小数 $0.p_1p_2\cdots$，则

$$\begin{array}{l} M:\{a_1, a_2, a_3, \cdots, a_n \cdots\} \\ \quad\ \ \updownarrow\ \ \ \updownarrow\ \ \ \updownarrow \qquad\ \ \updownarrow \\ N:\{1,\ 2,\ 3, \cdots,\ n, \cdots\}. \end{array}$$

若把 M 的一切元的小数表示式写出并依次排列，则有

$$\begin{array}{l} \overbrace{N}\ \ \overbrace{M} \\ 1 \leftrightarrow a_1 = 0.p_{11}p_{12}\cdots p_{1n}\cdots \\ 2 \leftrightarrow a_2 = 0.p_{21}p_{22}\cdots p_{2n}\cdots \\ 3 \leftrightarrow a_3 = 0.p_{31}p_{32}\cdots p_{3n}\cdots \\ \vdots \qquad\ \vdots \qquad\qquad\ \vdots \\ n \leftrightarrow a_n = 0.p_{n1}p_{n2}\cdots p_{nn}\cdots \\ \vdots \qquad\ \vdots \qquad \vdots \quad \vdots \quad \vdots \\ \underbrace{\ }\qquad \underbrace{\ } \end{array}$$

现用如下方法构造一个实数

$$q = 0.q_1q_2\cdots q_n\cdots。$$

如果选取 $q_1 \neq p_{11}, q_2 \neq p_{22}, \cdots, q_n \neq p_{nn}, \cdots$ 那么这个 $q \in [0,1]$，但数 q 和 a_i 中任何一个数都不相同，这是矛盾的，故[0,1]中的全体实数集 M 是不可数集，从而 R 亦是不可数集。如上构作新实数 q 并确证实数集 R 为不可数的过程便是康托的对角线方法的一种形式。

2. 计数原理及计数问题

集合的基数是从数量关系上刻画集合概念的本质特征的。研究集合间各种不同情况下的基数关系，是组合数

学中的重要内容。所谓计数原理指的就是处理组合数学中这类重要问题所遵循的思想、原则和方法。

(1) 映射观点下的排列组合和抽屉原理。

映射是揭示两集合间内在联系的重要手段，它是函数概念在抽象集合上的推广。用映射观点看待计数问题更易显示出计数问题的本质。

例如，从映射观点出发，排列组合数问题归结为两个集合间映射的数量关系问题。设

$$N_n = \{1,2,\cdots,n\}, N_r = \{1,2,\cdots,r\}。$$

则从集合 N_r 到 N_n 的所有单射的个数($r \leqslant n$)等于从 n 个各不相同的元素中，每次取出 r 个各不相同元素的所有不同排列的种数，且此种数为

$$P_n^r = \frac{n!}{(n-r)!}。$$

若记 $I(N_r,N_n)$ 为从集合 N_r 到 N_n 的所有单射的集合，则 P_n^r 就是集合 $I(N_r,N_n)$ 的元素个数，即

$$|I(N_r,N_n)| = P_n^r。$$

若从集合 N_r 到 N_n 的单射中，有两个单射 g、h 的像相同，则认为它们是同一组合。这样，从集合 N_r 到 N_n 的所有单射中按像相同定义等价关系，则其代表类集合中单射数目就是从 n 个不同元素中，每次取出 r 个不同元素的组合种数，且此种数为

$$C_n^r = \frac{n!}{r!(n-r)!}。$$

例如，设 $N_r = \{上,下\}$，$N_n = \{东,南,西\}$，所有单射为：

f_1:上 → 东，下 → 南；f_2:上 → 东，下 → 西；

f_3:上 → 南，下 → 东；f_4:上 → 南，下 → 西；

f_5:上 → 西,下 → 东;f_6:上 → 西,下 → 南。

$P_n^r = P_3^2 = 6$。但 f_1 与 f_3 其像都是{东,南},故相同。同理,f_2 与 f_5、f_4 与 f_6 分别相同;故可用$\{f_1,f_2,f_4\}$作为等价类,则 $C_n^2 = C_3^2 = 3$。

抽屉原理若用映射观点叙述,则为

设 f 是从 A 到 B 的任意一个映射,

$A = \{a_1,a_2,\cdots,a_n\}, B = \{b_1,b_2,\cdots,b_m\}(n > m)$,

且 $P = \left[\dfrac{n}{m}\right]$表示不小于$\dfrac{n}{m}$的最小整数,则必至少存在 $a_1,a_2,\cdots,a_p \in A$,使得

$$f(a_1) = f(a_2) = \cdots = f(a_p),$$

亦即集合 B 中至少存在一个元素其原像有 p 个。

抽屉原理又名鸽笼原理或重叠原理。这个原理的基础是分类,故按此原理求解问题时,首先要认清分类对象,找出分类规则,然后再去设计符合条件的抽屉。

(2) 容斥原理与加法原理。

1° 设有 n 个有限集合 $S_1,S_2,\cdots,S_n$,那么

$$|S_1 \cup S_2 \cup \cdots \cup S_n| = \sum_{1\leqslant i\leqslant n}|S_i| - \sum_{1\leqslant i<j\leqslant n}|S_i \cap S_j| + \sum_{1\leqslant i<j<k\leqslant n}|S_i \cap S_j \cap S_k| - \cdots + (-1)^{k-1}\sum_{1\leqslant i_1<i_2<\cdots<i_k\leqslant n}|S_{i1} \cap S_{i2} \cap \cdots \cap S_{ik}| + \cdots + (-1)^{n-1}|S_1 \cap S_2 \cap \cdots \cap S_n|$$

2° 设 $S_1,S_2\cdots,S_n$ 是有限集 I 的 n 个子集,则

$$|\overline{S_1} \cap \overline{S_2} \cap \cdots \cap \overline{S_n}| = |I| - |S_1 \cup S_2 \cup \cdots \cup S_n| = |I| - \sum_{1\leqslant i\leqslant n}|S_i| + \sum_{1\leqslant i<j\leqslant n}|S_i \cap S_j| - \sum_{1\leqslant i<j<k\leqslant n}|S_i \cap S_j \cap S_k| + \cdots + (-1)^k\sum_{1\leqslant i_1<i_2<\cdots<i_k\leqslant n}|S_{i_1} \cap S_{i_2} \cap \cdots \cap S_{i_k}| + \cdots + (-1)^n|S_1 \cap S_2 \cap \cdots \cap S_n|$$

其中符号 $\sum\limits_{1\leqslant i_1<i_2<\cdots<i_k\leqslant n}|S_{i_1} \cap S_{i_2} \cap \cdots \cap S_{i_k}|$ 表示 $S_1,\cdots$,

S_n 中任取 k 个集合的交的元素的个数的总和，$\overline{S_i}$ 是 S_i 对 I 的补集，即表示 I 中不属于 S_i 的元素组成的集合。

现对形式 1° 进行证明。至于形式 2° 不难由形式 1° 直接得到。对集合个数 n 进行数学归纳。

①$n=2$ 时，形式 1° 成为

$$|S_1\cup S_2|=|S_1|+|S_2|-|S_1\cap S_2|,$$

其正确性易于直接验证。

② 假设命题对 $n-1$ 成立，需证明命题对 n 也成立，注意到

$$S_1\cup\cdots\cup S_{n-1}\cup S_n=(S_1\cup\cdots\cup S_{n-1})\cup S_n,$$

由 ① 所得的公式可得

$$\begin{aligned}&|S_1\cup\cdots\cup S_{n-1}\cup S_n|\\=&|S_1\cup\cdots\cup S_{n-1}|+|S_n|-|(S_1\cup\cdots\cup S_{n-1})\\&\cap S_n|\\=&|S_1\cup\cdots\cup S_{n-1}|+|S_n|-|(S_1\cap S_n)\cup\cdots\cup\\&(S_{n-1}\cap S_n)|\qquad(*)\end{aligned}$$

由归纳假设，对于 $n-1$ 个集合 $S_1\cap S_n,\cdots,S_{n-1}\cap S_n$，必有

$$\begin{aligned}&|S_1\cap S_n)\cup(S_2\cap S_n)\cup\cdots\cup(S_{n-1}\cap S_n)|\\&=|S_1\cap S_n|+|S_2\cap S_n|+\cdots+|S_{n-1}\cap S_n|\\&-|(S_1\cap S_n)\cap(S_2\cap S_n)|-|(S_1\cap S_n)\cap(S_3\cap S_n)|\\&-\cdots-|(S_{n-2}\cap S_n)\cap(S_{n-1}\cap S_n)|+\cdots+(-1)^{n-2}\\&|(S_1\cap S_n)\cap(S_2\cap S_n)\cap\cdots\cap(S_{n-1}\cap S_n)|\\&=|S_1\cap S_n|+|S_2\cap S_n|+\cdots+|S_{n-1}\cap S_n|-|S_1\\&\cap S_2\cap S_n|-|S_1\cap S_3\cap S_n|-\cdots-|S_{n-2}\cap S_{n-1}\cap\\&S_n|+\cdots+(-1)^{n-2}|S_1\cap S_2\cap\cdots\cap S_n|\qquad(**)\end{aligned}$$

又由归纳假设，对于 $n-1$ 个集合 $S_1,\cdots,S_{n-1}$ 必有

$$|S_1 \cup S_2 \cup \cdots \cup S_{n-1}|$$
$$= |S_1| + |S_2| + \cdots + |S_{n-1}| - |S_1 \cap S_2| - |S_1 \cap S_3| \cdots - |S_1 \cap S_{n-1}| + |S_1 \cap S_2 \cap S_3| + \cdots + |S_{n-3} \cap S_{n-2} \cap S_{n-1}| - \cdots + (-1)^{n-2}|S_1 \cap S_2 \cap \cdots \cap S_{n-1}| \quad (***)$$

再把(**),(***)两式代入(*),经整理,即得形式1°的容斥原理。

容斥原理的特例是加法原理,这只要注意到:

形式1°中$S_1, S_2, \cdots, S_n$两两不交,即$S_i \cap S_j = \varnothing (i \neq j)(i, j = 1, 2, \cdots, n)$,这时,形式1°为

$$|S_1 \cup S_2 \cup \cdots \cup S_n| = |S_1| + |S_2| + \cdots + |S_n|$$

即是加法原理。这时,$S_1, S_2, \cdots, S_n$是$S = \bigcup_{i=1}^{n} S_i$的一个分划,亦即容斥原理是加法原理的推广。

容斥原理又称逐步淘汰原理。1920年挪威数学家布龙(V. Brun)将容斥原理运用于数论,作为一种筛法去确定素数的个数。例如,要数11到30内素数的个数,这区间共20个数,先数出其中合数的个数。例如数2、3、5的倍数,即

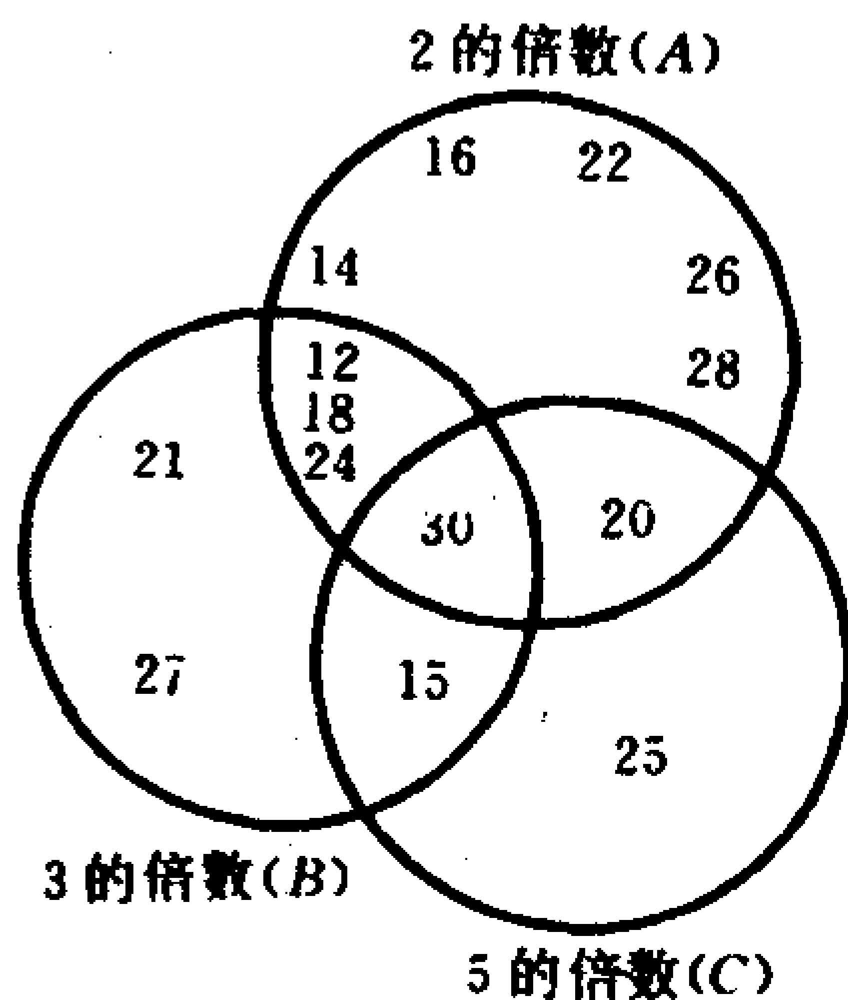


图3.2

$$|A| + |B| + |C| = 10 + 7 + 4 = 21。 \quad ①$$

由于A、B、C有相同部分,计数重了,需减去

$$|A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| = 21 - 4 - 2 - 2 = 13。 \quad ②$$

减多了,需要补加上

$|A|+|B|+|C|-|A\cap B|-|B\cap C|-|A\cap C|+|A\cap B\cap C|=13+1=14$。 ③

故素数的个数为 $20-14=6$ 个。

由上述分析知，③ 是精确的，①、② 分别太大和太小，布龙的诀窍在于不用 ③，而用 ① 和 ② 来代替，因为尽管 ① 和 ② 不正确，但比 ③ 简单，它能给出合理的近似。布龙的这一思想，具有重要的意义，尤其是求区间段数目很大的素数个数时更是如此。

例 3 由数字 1、2 和 3 组成 n 位数，要求 n 位数中 1，2 和 3 的每一个至少出现 1 次，求所有这种 n 位数的个数。(匈牙利数学竞赛试题)

解：记由数字 1、2 和 3 组成的 n 位数全体构成的集合为 S，则 $|S|=3^n$。又 S 中不含 1 的 n 位数的集合记作 A_1，A_2、A_3 与 A_1 同样理解。于是 $\overline{A_1}\cap\overline{A_2}\cap\overline{A_3}$ 即是 S 中同时含有数字 1、2 和 3 的 n 位数全体构成的集合，要求的即是这一集合的元素个数。由容斥原理，知

$$|\overline{A_1}\cap\overline{A_2}\cap\overline{A_3}|$$
$$=|S|-|A_1|-|A_2|-|A_3|+|A_1\cap A_2|+|A_1\cap A_3|+|A_2\cap A_3|-|A_1\cap A_2\cap A_3|。$$

由于 A_i 是 S 中所有不含数字 i 的 n 位数的集合，故 $|A_i|=2^n(i=1,2,3)$，而 $A_i\cap A_j$ 是所有不含数字 i 和 j 的 n 位数集合，所以

$$|A_i\cap A_j|=1,1\leqslant i<j\leqslant 3。$$

显然

$$A_1\cap A_2\cap A_3=\varnothing，而 |A_1\cap A_2\cap A_3|=0，$$

故 $$|\overline{A_1}\cap\overline{A_2}\cap\overline{A_3}|=3^n-3\times 2^n+3。$$

即由数字 1、2、3 组成的三个数字每个至少出现一次的 n 位数的个数为 $3^n-3\times 2^n+3$。

本问题求解的关键在于构造集合，尔后把原问题化为求解有关集合的元素的个数问题。

例4 在1至100的整数中任取51个数，其中必有一数是另一数的倍数。

解：依题意，可设法把1—100间的整数分成50类，即50个抽屉。这样，任取51个数，至少有2个数属于同类。鉴于此，在分类时，只要使同类中的数有倍数关系即可。因1—100这100个数中，刚好50个是奇数，50个是偶数，而每个偶数总可写成：奇数 $\times 2^h$（h是不大于6的正整数）的形式，故可按50个奇数作基础，把每个奇数和它的2^h倍的诸偶数（如果有的话）归入同一类。即

$M_1 = \{1, 1\times 2, 1\times 2^2, \cdots, 1\times 2^6\}$；

$M_2 = \{3, 3\times 2, 3\times 2^2, \cdots, 3\times 2^5\}$；

$M_3 = \{5, 5\times 2, 5\times 2^2, \cdots, 5\times 2^4\}$；

$M_4 = \{7, 7\times 2, 7\times 2^2, \cdots, 7\times 2^3\}$；

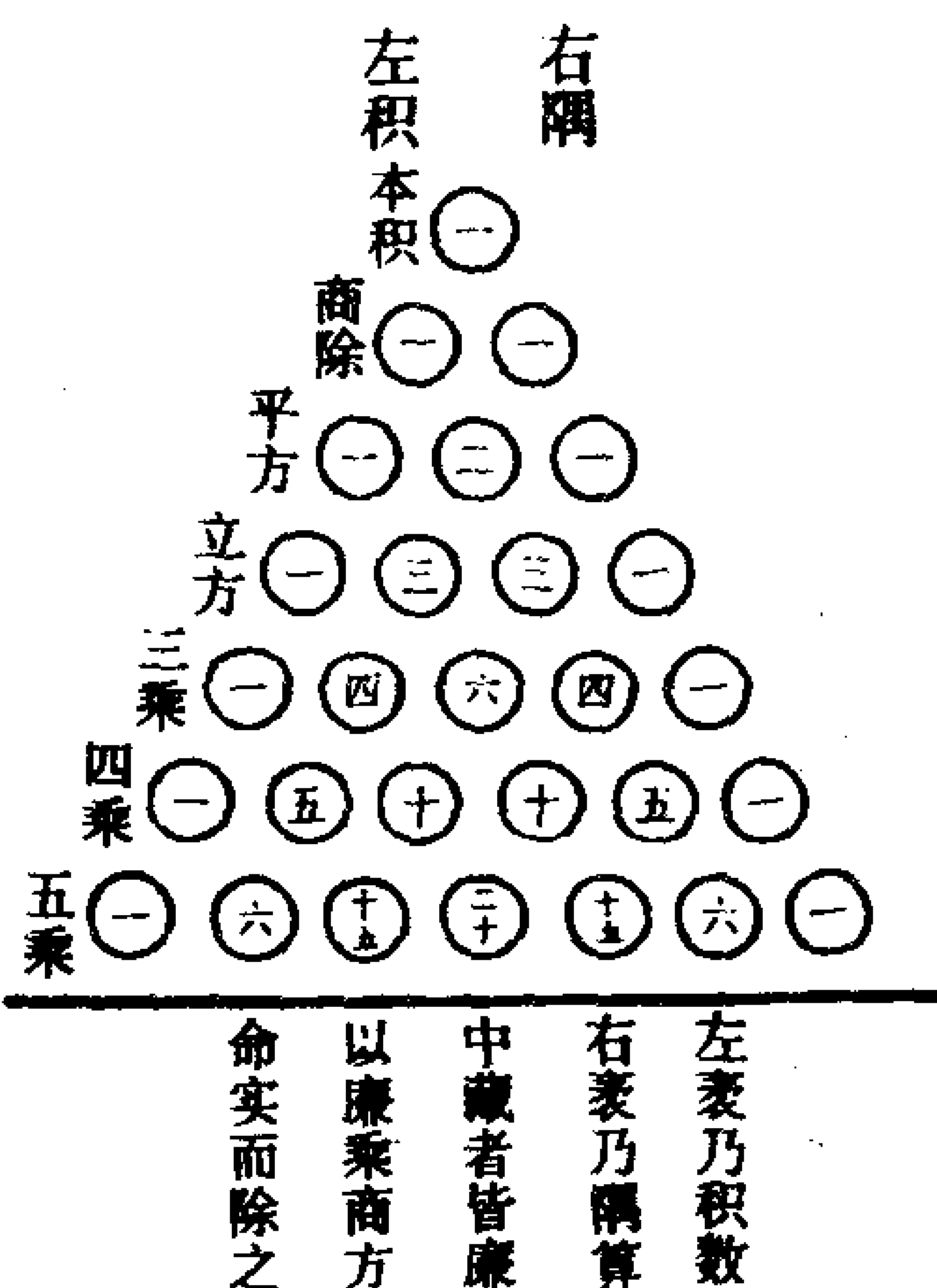
……

$M_{49}\{97\}$；

$M_{50} = \{99\}$。

这样共50类。这50类中元素个数的和为100，由抽屉原理，在任取的51个数中必至少有两个数属于同类，从而必有一数是另一数的倍数。

我国古代在计数原理方面，著名的有北宋时贾宪的二项系数表即“开方作法本源”图。可惜贾宪之书已失传，但杨辉《详解九章算法》(1261) 记有：“出释锁算书，贾宪用此术。”证明杨辉是征引了贾宪的材料。它实际是一个指数为正整数的二项式定理的系数表，从图3.3下面注释的前三句知，每一层数字表示$(a+x)^n$展开式中各项的系数。最外边左、右斜线上的数字，分别是“积”a^n



(1) 开方作法本原图

图 3.3

的系数，和"隅"x^n 的系数；中间的许多数字如 ㊁、㊂㊂、㊃㊅㊃ 等分别是各项"廉"$a^{n-r}x^r$ 的系数($1\leqslant r<n$)。"贾宪三角"之所以称为"开方作法本源图"是因为它是我国古代用来开方的工具。朱世杰的"古法七乘方图"(图3.3)比贾宪的开方作法本源图多两层，并增添了斜线用以表明每行各个系数与其上一层系数存在的关系。

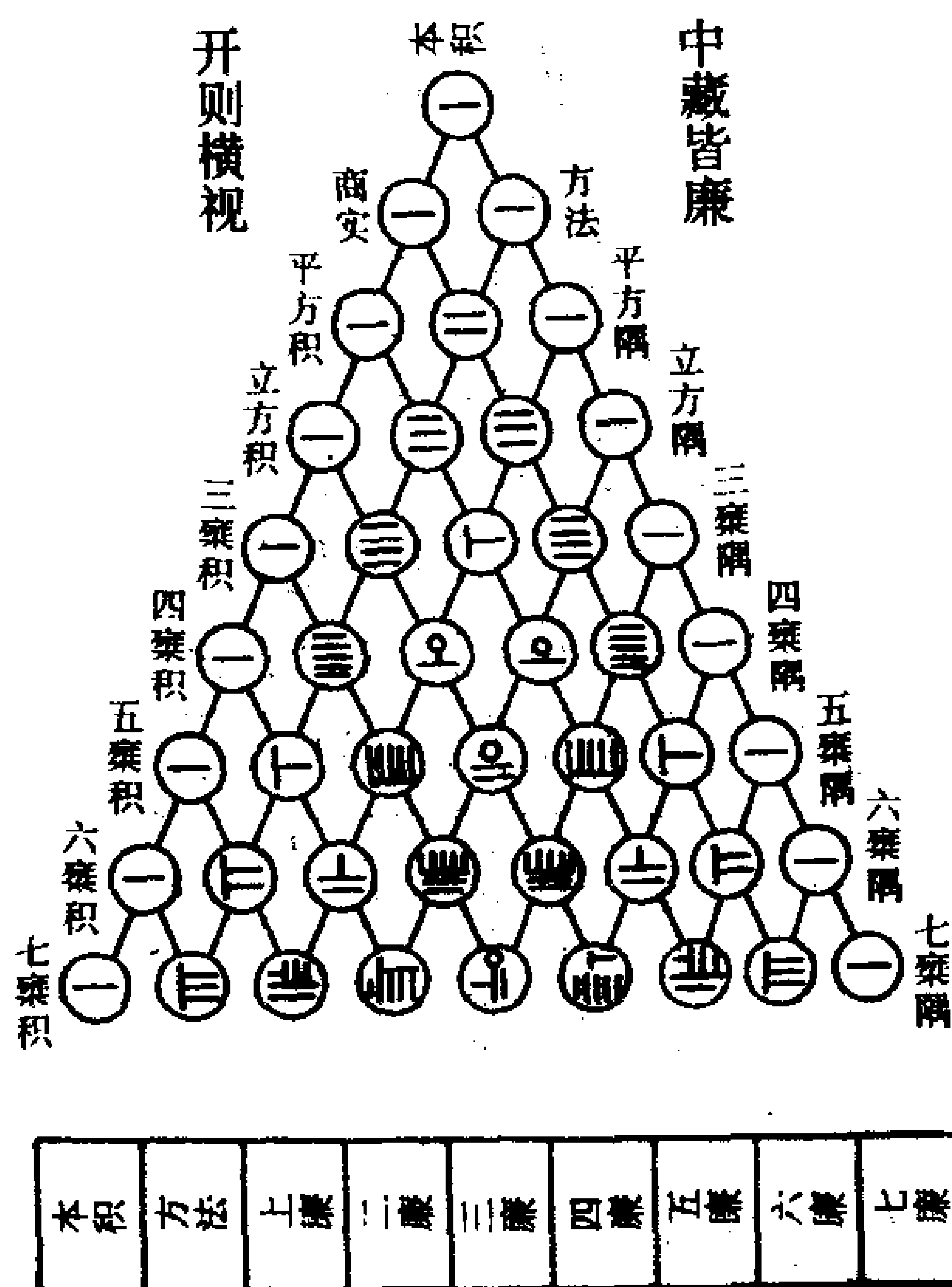
例如，第七层表示

$$(x+a)^6=x^6+6ax^5+15a^2x^4+20a^3x^3+15a^4x^2+6a^5x+a^6,$$

其第二项系数是上层第一、二项系数 1 与 5 之和。

"贾宪三角"的提出不迟于1200年，而欧洲巴斯加在1654 年才提出。

此外，朱世杰还得出了组合恒等式：



(2) 古法七乘方图

图 3.3

$\sum\limits_{1\leqslant r\leqslant n}\binom{r+p-1}{p}=\binom{n+p}{p+1}\ (1\leqslant p\leqslant 6)$，符号 $\binom{r+p-1}{p}$ 即 C^{p}_{r+p-1}。

清代李善兰证明了此恒等式对一切正整数 p 成立，这是我国数学家为组合数学作出的重要贡献。

3.2 位值原则与记数方法

1. 位值原则与数的层次结构

数学思维中被认为是“最美妙的发明之一”的位值制记数法，是指选取某个数作为基数，并用不同符号分别表示比基数小的各数，通过对符号的不同排列来表示出

任意数的一种方法。若基数为 g，则所表示的数称为 g 进位制的数。在位值制记数法中，数码代表的值，不仅与数码的符号有关，而且与数的进位制度及其符号在数中所处的位置有关，这种确定数符号在数中代表的值的原则称为位值原则。位值原则的诀窍在于不同单位的数，采用了相同的符号，因而如同丹齐克所说的，它是"方法上的根本变革"。

例如，$g=8$。用 0，1，…，7 等八个数码代表 8 进制数，则 $(3736)_8$ 中的 6、3、7、3 分别代表 6×8^0，3×8^1，7×8^2，3×8^3。其对应的 10 进制数为

$$\begin{aligned}(3736)_8 &= 3\times 8^3+7\times 8^2+3\times 8^1+6\times 8^0\\ &= (2014)_{10}。\end{aligned}$$

同一个 3，在不同的位置代表不同的值。

又如，$(110101.1101)_2=2^5+2^4+2^2+2^0+2^{-1}+2^{-2}+2^{-4}=(53.8125)_{10}$，$(AE7)_{16}=10\times 16^2+14\times 16^1+7\times 16^0=(2791)_{10}$

其中，$A\sim F$ 表示 16 进制中 $10\sim 15$ 六个基数。

对于任意 g 进位数间的四则运算，在遵循"逢 g 进一"的原则下完全类同于普通十进位制数间的运算，对于不同进位制之间的运算，要化为同一进位制后才能进行。例如，

$(2.021)_3\times(1.2)_3$

$=(10.2022)_3$，

而求 $(25)_{10}-(10.11)_2$ 则要化为同是二进制或同是十进制以后再进行，如将 $(10.11)_2$ 化为十进的，有

```
      2.021
 ×)     1.2
 ----------
    11 112
    20 21
 ----------
  10.2 022
```

$$(10.11)_2=1\times 2^1+1\times 2^{-1}+1\times 2^{-2}=(2.75)_{10}。$$

从而

$$(25)_{10}-(10.11)_2=(25)_{10}-(2.75)_{10}$$
$$=(22.25)_{10}。$$

又将一个十进制的正整数化为 g 进的数，只要用基数 g 对这个数不断地作带余除法直到商为零为止，然后按相反的顺序列出各余数即可。如$(25)_{10}$化为二进的，有

$$\begin{array}{r|rl}
2 & 25 & \\
2 & 12 & 余\ 1 \\
2 & 6 & \quad 0 \\
2 & 3 & \quad 0 \\
2 & 1 & \quad 1 \\
 & 0 & \quad 1
\end{array}\quad\uparrow$$

则 $(25)_{10}=(11001)_2$。

所以

$$(25)_{10}-(10.11)_2=(11001)_2-(10.11)_2$$
$$=(10110.01)_2$$

显然 $(10110.01)_2=(22.25)_{10}$。

故两种不同方法运算是等效的。

位值的记数方法，通过有限个数符号，借助于位值原则，表达了一切的数，确是个了不起的创造。18至19世纪法兰西著名数学家拉普拉斯评论道："用十个记号来表示一切的数，每个记号不但有绝对的值，而且有位置的值，这种巧妙的方法出自印度。这是一个深远而又重要的思想，它今天看来如此简单，以致我们忽视了它的真正伟绩。但恰恰是它的简单性以及对一切计算都提供了极大的方便，才使我们的算术在一切有用的发明中列在首位；而当我们想到它竟逃过了古代最伟大的两位人物阿基米德和阿波罗尼的天才思想的关注时，我们更感到这成就的伟大了。"[26] 著名科学史家李约瑟也称道说："如果没

有这种十进制，就几乎不可能出现我们现在这个统一化的世界了。"[27]

记数法原理与自然界中物质的分层结构原理颇为相似。自然界中的物质结构系统具有层次性，且这种层次结构是多样的，统一的。即是说，任何一个结构层次一方面具有自身的特点，另一方面它无论在时间序列上、还是在空间广延上，又都只是系统中的一个关节点，是特定的时空形式的统一。层次之间、系统之间不仅互相联系、相互包摄，而且纵横交错。在记数法中，由于基数的不同，带来了层次结构的深度（层次数）、宽度（每层元素数）的不同。基数大的，宽度大，深度小，因而容量亦大。而基数小的，宽度小，深度大，容量就小。例如，十进制的宽度为10，一位十进制数最多相当于四位二进数（$9_{10} = 1001_2$），故二进制的容量最小。但由于这种数仅有两种符号，便于借用通电、断电加以表达而利于自动计算，故被电子计算机广泛采用。

事物的分层带来了事物结构的简单性，功能的有效性和变化的灵活性等。现代科学管理中这种分层结构原理的应用产生了层次管理和管理的宽度和深度等概念。

2. 中国古代对位值制的贡献

如上述拉普拉斯关于位值制的论断，不失为至理名言。遗憾的是，它忽略了中国古代对位值制的重大贡献。

我国古代早有原始记数的文字记载。《易传·系辞》说："上古结绳而治，后世圣人易之以书契。"李迪在《中国数学史简编》[28] 中展示了少数民族记数法的许多文物，马家窑文化等遗址中出土的新石器时代的彩陶钵口沿上50多种刻画符号都为这些记数法提供了充足的例证。原始记数的发展便是数码或筹码记数，这是算筹计数

的摹写，早见于战国时代的币货币。64 记为𝍮𝍣(《古钱大辞典》)。1978 年在河南登封县出土的战国早期陶文中也有将 41、52 分别记为𝍬𝍠、𝍭𝍡，其记数方法是位值的。

记数的位值方法的完整记载，见于《夏侯阳算经》(实为韩延《算术》，8 世纪下半叶成书)。书中写道："一纵十横，百立千僵。千、十相望，万、百相当。满六以上，五在上方。六不积算，五不单张。"1 至 9 的数码，纵式为 𝍠 𝍡 𝍢 𝍣 𝍤 𝍥 𝍦 𝍧 𝍨，横式为 𝍩 𝍪 𝍫 𝍬 𝍭 𝍮 𝍯 𝍰 𝍱。数零，始则用空位表示，继则用 □，后来演化为 ○。如敦煌石室发现的唐写卷子本《立成算经》的"九九表"中，三六一十八的"18"写为"𝍩 𝍧"，"通前百八"的"108"为"𝍠　𝍧"，零用空位表示。宋蔡沈《律吕新书》(12 世纪) 卷一"律吕本原"内记有"丑林钟十一万八千 □□ 九十八"，零用 □ 表示，而金大明历(1180) 中，已有"初损四百 ○ 三"、"益五百 ○ 五"的记载，零用 ○ 表示。九个筹码中，𝍣后来简为 ×，𝍤在十位、千位为♁，单、百、万位为 ○̄，𝍨在十、千位为 ⋆，个、百、万位为 ×̄。许多时候，繁简形式的数码常混合使用，即使同一问题中亦如此。如秦九韶《数书九章》"推计互易"问题中共壳 175938(升)、得米 39600(升)分别记为𝍩 𝍦 ♁ ×̄ 𝍫 𝍧(升) 和𝍢 ⋆ 𝍥 ○○(升)，而出壳 92400(升)、踏麹 30294(斤) 则记为𝍨 𝍪 𝍣 ○○(升) 和𝍢 ○ 𝍡 𝍱 𝍣(斤)。

我国的十进制位值记数法，既优于古巴比伦 60 进位记数法，又优于古希腊、罗马的十进非位值记数法，是当时世界上最先进、最方便的记数方法。筹算中除法按商、实(被除数)、法(除数) 自上而下排列，实际就是带分数的竖写表示，比印度的类似表示法最少早 4 个世纪；小数记法，赝本《夏侯阳算经》常以某一整数单位表示数值，

不列出以下的名数单位，1525.9375 匹记成 1525 匹 9375，名数匹兼起小数点的作用，比西方斯蒂文(Stevin，1585)的记法 1525⓪9①3②7③5④ 简洁，且早几个世纪；负数的概念及表示也“超出了其他国家几世纪之久”。

我国位值记数除十进制外，还有二进制的。《周易》中阳爻、阴爻符号“—”和“－－”每三个组成一卦叫八卦，如 ☰ 为乾，☷ 为坤，☵ 为坎，☲ 为离等。八卦的每两个相重叠(可重复)得 $8^2=64$ 种符号，代表 64 种“别卦”，这实际就是以“—”和“－－”为基本记数符号的 64 个不同的二进制数符号，曾被莱布尼兹誉为“是流传在宇宙间科学中的最古的纪念物”[21]。我国对位值制的贡献，完全如李约瑟所论断的“在西方后来所习见的‘印度数字’的背后，位值制早已在中国存在了 2000 年”[27]。

3. 位值制的应用

位值原则和各种进位制一直被人们广泛使用着。有人曾对美洲——印第安之间的几百个部落进行研究，发现三分之一的部落用十进位制，百分之十的用 20 进位制，而利用三进位的则少于百分之一。现代随着电子计算机的普遍使用，进位制尤其是二进位制应用更为广泛，以下展示的是几个具体应用的例子。

例 1 用二进制原理构造数学问题。

设有若干张表格，只要你说出自己的年龄数所在的表格号码，则立即可猜到你的年龄。这是因为表格是按下述方法制作的：

将有关十进制数化为二进制数。设二进制数中的 0 表示对应的十进制数不填入表中，0 在第几位(自左至右)的，第几表中就不填入数。1 表示要将对应的十进制数填入表中，1 在第几位的，就将数填入第几表中。例如，

$(000001)_2=(1)_{10}$，表示 1 只填入第 1 表中，

$(000010)_2=(2)_{10}$，则 2 只填入第 2 表中，

$(000011)_2=(3)_{10}$，3 只填入第 1、2 表中，

……

$(111111)_2=(63)_{10}$，63 要填入第 1—6 的每一表中。

若某人自己说自己的年龄在第 1、3、4 表中，则由

$(001101)_2=(13)_{10}$

立即可说出他的年龄数是 13。这就是猜年龄的奥秘。

与猜年龄应用的原理相似，有如下的问题：

63 箱外形完全一样的货物，其中 62 箱正品，1 箱副品，已知副品比正品轻，是否有办法尽量减少秤的次数，又能把副品的一箱找出来？

问题的答案是，只需秤 6 次。办法是把 63 箱依次编号为 1，2，…，63，并用二进制表示。每数需写成 6 位数，即 $(000001)_2$，$(000010)_2$，…，$(111111)_2$。现把每位中出现"1"的箱号归在一起，于是就得 6 个组，然后把每个组中的箱子集中起来秤一次，若重量正常，记为 0，轻了则记为 1，这样共秤 6 次，即得一个 6 位的二进制数，它就是副品的箱号数。这里，副品的箱号数相当于猜年龄中的年龄数，重量为正常，相当于年龄问题中年龄数不在表中，故记为 0，不正常的记为 1。因此，6 次所得的二进制数就是副品的箱号数。

例 2　取胜的对策。

设有 18 根火柴分成 3 堆，分别有 4、5、9 根，两人轮流从三堆中取，每次至少一根，至多一堆，但不允许同时在两堆中取，取得最后一根火柴者为胜，问先取者应如何取才能稳操胜券。

将三堆火柴的根数分别用二进制表示为 $(100)_2$，

$(101)_2$，$(1001)_2$。再将它们的各位数字分别相加但不进位，得$(1202)_2$，先取者想获胜，必须把$(1202)_2$中各位数全变为偶数，现$(1202)_2$仅有一个奇数1，于是在9根的一堆中取走8根，这时剩下$(100)_2$，$(101)_2$，$(001)_2$。其三数和的各位数“全是偶数”，这样，无论第二人怎样取，所取的那一堆火柴的根数(用二进制表示)至少有一位上数字发生了变化，从而使“全是偶数”变为“不全是偶数”。比如第二人在第一堆中取了2根，剩下$(010)_2$，$(101)_2$，$(001)_2$。这时三数和$(112)_2$各位不全是偶数。于是，先取者需设法在某一堆中取走若干根，使剩下根数之和的各位数全是偶数，故他应在第二堆中取2根，这时剩下的为$(010)_2$，$(011)_2$，$(001)_2$。如此下去，先取者自始至终创造了“全是偶数”的局面，并至最后出现$(0)_2$为止，从而保证了先取者必取走最后一根火柴而获胜。

以上的方法适用于一般情况。若有4堆火柴，根数为2，3，6，6，游戏规则仍不变。因为

$(2)_{10}=(010)_2$，$(3)_{10}=(011)_2$，$(6)_{10}=(110)_2$，$(6)_{10}=(110)_2$。二进制数各位数字之和为2，4，1，不是全偶数，故先取者可在第二堆中取走一根，变为2，2，6，6，便掌握了胜利局势。

3.3　等价原理与大衍求一术

1. 等价原理与同余

等价关系是指满足反射律($a\sim a$)、对称律($a\sim b\Rightarrow b\sim a$)、传递律($a\sim b\ b\sim c\Rightarrow a\sim c$)的关系，这是研究数学结构的一种基本关系。一个等价关系确定系统的一种分类，因而可以从等价关系研究事物的相同与相异，同类

与异类，这种类就是等价类。等价类的思想实质是舍弃了对象之间的某些区别，而把具有某种特征的相异对象集中起来，并在等价意义下，把它们当作“类似”的，甚至是“等同”的。例如，全等的图形，不管其所在的位置怎样不同，都是等价的，称为合同等价。而形状相似的图形都称为是相似等价的。等价关系是人们认识和研究事物的眼光和途径。它是转换人们研究方法的理论基础，即通过等价进行转换。

同余也是一种等价。所谓a、b两数关于模n同余是指a和b被$n(>0)$除后所得的余数相等，记为

$$a \equiv b \pmod{n}。$$

对同余，显然有

1° $a \equiv a \pmod{n}$；

2° 若$a \equiv b \pmod{n}$，则$b \equiv a \pmod{n}$；

3° 若$a \equiv b \pmod{n}$，$b \equiv c \pmod{n}$，则$a \equiv c \pmod{n}$。

故同余关系是一种等价关系。所有关于模n同余的元素组成的等价类称为剩余类。剩余类中的每一元素都可作为类的代表，当选取其最为简单的元素组成代表集合时，对原集合的研究就可转化为对代表集合的研究。

例如，整数集Z，关于模n，分为n个等价类：

$[0]_n, [1]_n, \cdots, [n-1]_n$。

当$n=2$时，有等价类0类和1类，即$[0]_2$、$[1]_2$，分别就是偶数集和奇数集。若将n个等价类

$$[0]_n, [1]_n, \cdots, [n-1]_n$$

作为元素，组成集合

$$Z_n = \{[0], [1], \cdots, [n-1]\}$$

记做$Z|(\bmod n)$或$Z|R$（R是同余关系），称做以n为模的商集。显然，对于

$$Z_n = Z|(\text{mod } n) = \{[0],[1],\cdots,[n-1]\},$$

其代表集合为$\{0,1,\cdots,n-1\}$,即是非负最小完全剩余系。在处理实际问题中,把将对 Z 的考察,化为对非负最小完全剩余系的考察,常可使问题得到简化。

例如,今天是星期六,10^{1000} 天之后是星期几?

因 10^{1000} 数目太大,不能直接进行研究。注意到,

$$10 \equiv 3(\text{mod } 7), 10^2 \equiv 3^2 \equiv 2(\text{mod } 7),$$

则 $10^3 \equiv 6 \equiv (-1)\ (\text{mod}7)$,从而 $10^6 \equiv 1(\text{mod}7)$,

所以,

$$10^{1000} = 10^{6\times166+4} = (10^6)^{166}\cdot10^4 \equiv 1^{166}\cdot4(\text{mod}7)$$
$$\equiv 4(\text{mod}7)。$$

即 10^{1000} 被 7 除余 4,故 10^{1000} 天之后应是星期三。

求解本题的要点是在同余条件下将对

$$10^{1000} = 10^{6\times144+4} = (10^6)^{144}\cdot10^4$$

的考察化为对 10^6 与 10^4 的考察,而这种考察是通过建立它们与代表集合$\{0,\cdots,6\}$间的关系来实现的。

2. 物不知数与中国剩余定理

中算家关于同余思想和一次同余问题的研究源远流长,西汉历算家推算上元积年便开始了这方面的探索。中国数理天文学的早期发达与历法的不断改革,使得治历演纪成为中国天算史上一个经常性的课题,而这些课题的不断探索,使中算家对同余问题的研究成果达到了世界数学的最高水平。

载于《孙子算经》的物不知数问题,实际就是作者依据当时天文学家推算"上元积年"的算法经简化后编写出来的。"物不知数"的解法后来被欧洲誉为"中国剩余定理"。

《孙子算经》的作者与成书年代均不可考,但据《隋

书·经籍志》中录有孙子算经二卷，推算其成书至晚在4世纪之前。物不知数问题是：

“今有物不知其数，三三数之剩二，五五数之剩三，七七数之剩二，问物几何？”

问题后还有答与术。

“答曰：二十三。”

“术曰：三三数之剩二，置一百四十，五五数之剩三，置六十三，七七数之剩二，置三十。并之，得二百三十三，以二百一十减之，即得。”

“凡三三数之剩一，则置七十，五五数之剩一，则置二十一，七七数之剩一，则置十五。一百六以上，以一百五减之，即得。”

术的第二段，“凡三三数之剩一，则置七十”意谓取70为基数，因为用3除它余1，同时它又分别是5、7的倍数。

同理，“五五数之”、“七七数之”的二句，是指取21、15为基数，用5、7分别除它们都余1，且21、15又分别是3、7和3、5的倍数。故将三基数相加得106减去3、5、7三数的最小公倍数105后其余仍不变。

第一段术文“三三数之剩二，置一百四十”是指140是用3除余2，同时又能被5、7整除的数，这是因为70是用3除余1，同时又能被5、7整除的数。同理，63是用5除余3，又可被3、7整除的数，30是用7除余2，同时又能被3、5整除的数。则三数并之得233，减去150的2倍210得23仍是用3、5、7除余数分别为2、3、2的数，故23即所求的数。

“物不知数”中的术具有一般性，其一般形式即是孙子定理，后被誉为中国剩余定理，用现代数学语言叙述

是：

若某数 x 分别被 $m_1,\cdots,m_n$ 除所得的余数为 $u_1,\cdots,u_n$，则 x 可表示为

$$x = k_1u_1 + k_2u_2 + \cdots + k_nu_n + k \cdot D,$$

其中 k_1 是 $m_2,m_3,\cdots,m_n$ 的公倍数，而且被 m_1 除得余数为 1；k_2 是 $m_1,m_3,\cdots,m_n$ 的公倍数，而且被 m_2 除得余数为 1；$\cdots$，k_n 是 $m_1,m_2,\cdots,m_{n-1}$ 的公倍数，而且被 m_n 除得余数为 1。D 是 $m_1,m_2,\cdots,m_n$ 的最小公倍数，$m_1,m_2,\cdots,m_n$ 两两互质，k 是任意整数，可根据实际需要而决定。

在“物不知数”题中，$m_1=3,m_2=5,m_3=7,u_1=2,u_2=3,u_3=2$。由术知 $k_1=70,k_2=21,k_3=15,D=105$。则

$$\begin{aligned} x &= 70 \times 2 + 21 \times 3 + 15 \times 2 + 105k \\ &= 233 + 105k。\end{aligned}$$

取 $k=-2,x=233-105 \times 2=23$。

“物不知数”解法的传播始于英人伟烈亚力写入他的《中国数学笔记》中，并于1852、1853年先后在“华北前锋报”和“上海年鉴”发表，不久传入欧洲。

3. 余米推数与大衍求一术

解决《孙子算经》“物不知数”的关键是求 k_i，秦九韶在《数书九章》(1247)中提供的方法称为“大衍求一”术，其求法不仅简单、明了，而且还适用于除数并非两两互质的情况。为简单起见，仍以除数两两互质的情况为例。

设有 n 个两两互质的正整数：$m_1,\cdots,m_n$，又设 $u_1,\cdots,u_n$ 为已知整数，则有唯一的一个整数 $u \geqslant 0$，且 $u < m = m_1m_2\cdots m_n$，使

$$u \equiv u_j \pmod{m_j}, 1 \leqslant j \leqslant n。$$

秦九韶在大衍求一术中，称 m_j 为定母，m 为衍母，u_j 为余

数，并称m_j除m/m_j之余为奇数（意谓零头数）。则所求的u为

$$u \equiv \Sigma u_j k_j (m/m_j) (\mathrm{mod}\ m),$$

其中k_j秦氏称为乘率，(m/m_j)称为衍数。

以下就秦九韶《数书九章》中"余米推数"问题为例来说明大衍求一术的具体实施方法。原题照录如下：

"问有米铺，诉被盗去米一般三箩，皆适满，不记细数。今左壁箩剩一合，中间箩剩一升四合，右壁箩剩一合，后获贼，系甲乙丙三名。甲称当夜摸得马杓，在左壁箩，满舀入布袋，乙称踢着木履，在中箩，舀入袋，丙称摸得漆椀，在右边箩，舀入袋，将归食用，日久不知数。索到三器，马杓满容一升九合，木履容一升七合，漆椀容一升二合。欲知所失米数，计赃结断三盗各几何。"

今以合为单位，则依题意：

剩米数 = 余数 $u_j = 1, 14, 1$，

盗器容量 = 定母 $m_j = 19, 17, 12$，

衍母 $m = m_1 m_2 m_3 = 19 \times 17 \times 12 = 3876$，

衍数 $m/m_j = 204, 228, 323$，

奇数 = m_j 除 m/m_j 之余 $= 14, 7, 11$，

秦九韶求乘率k_j的方法是："置奇右上，定居右下，立天元一于左上，先以右上除右下，所得商数与左上一相生入左下，然后乃以右行上下，以少除多，递互除之，所得商数，随即递互累乘，归左行上下，须使右上末后奇一而止，乃验左上所得，以为乘率。"这里，"奇"、"定"分别指"奇数"和"定数"，运算前应置于右上和右下。"立天元一"，莫绍揆先生以为是指未知数x的系数，记于左上。将这三个数字写成方阵，然后按上述术文逐步变换，使右上数减小到1，此时左上数便是所求的乘率。如k_1的求法

是：

$$\boxed{\begin{matrix}1 & 14\\ & 19\end{matrix}}\to\boxed{\begin{matrix}1 & 14\\ & 5\end{matrix}}\,1\to\boxed{\begin{matrix}1 & 14\\ 1 & 5\end{matrix}}\to\boxed{\begin{matrix}1 & 4\\ 1 & 5\end{matrix}}\,2\to\boxed{\begin{matrix}3 & 4\\ 1 & 5\end{matrix}}\to\boxed{\begin{matrix}3 & 4\\ 1 & 1\end{matrix}}\,1\to$$

$$\boxed{\begin{matrix}3 & 4\\ 4 & 1\end{matrix}}\to\boxed{\begin{matrix}3 & 1\\ 4 & 1\end{matrix}}\,3\to\boxed{\begin{matrix}15 & 1\\ 4 & 1\end{matrix}}$$

于是得 $k_1 = 15$。求 k_2 的算式是：

$$\boxed{\begin{matrix}1 & 7\\ & 17\end{matrix}}\to\boxed{\begin{matrix}1 & 7\\ & 3\end{matrix}}\,2\to\boxed{\begin{matrix}1 & 7\\ 2 & 3\end{matrix}}\to\boxed{\begin{matrix}1 & 1\\ 2 & 3\end{matrix}}\,2\to\boxed{\begin{matrix}5 & 1\\ 2 & 3\end{matrix}}$$

即 $k_2 = 5$。

同理求得 $k_3 = 11$。

$u \equiv \Sigma u_j k_j (m/m_j) \pmod{m}$

$\equiv [1 \times 15 \times 204 + 14 \times 5 \times 228 + 1 \times 11 \times 323] \pmod{3876}$

$\equiv 22573 \pmod{3876}$　　所以 $u = 3193$。

在秦九韶的《数书九章》中，处理一次同余问题的方法有二，其一是“大衍总数术”，其二是“治历演纪术”。前者是一次同余式组问题的一般解法，后者是专为历元推算设计的演算程序；二者皆基于“大衍求一术”。秦氏的大衍总数术将孙子剩余定理推广到了模数两两并非互质的一般情况。他设计的算法要比近代析素因子法在实际计算上有较大的优越性。在欧洲，最早接触一次同余问题的是意大利数学家斐波纳契，他在 1202 年写的《算盘之书》中给出了两个一次同余问题，但没有一般的算法。直至十八、九世纪才由大数学家欧拉(1743)、高斯(1801) 各自重新获得与“孙子定理”相同的结论，并就模数两两互质的情况做出严格的证明。

4. Hill 密码

研究秘密信息的编码和译码的学科叫做密码学。同余在密码学中也有重要的应用。

密码学中，代码叫密码，未编码的信息叫做明文，译成代码的信息叫做密文。由明文转换成密文的过程叫做编码。由密文转换成明文的过程叫做译码。最简单的密码是代用码，它是将字母表中的每一字母用不同的字母来代替的。由于它保持着各个字母出现的频率，故使它易于为统计方法所破译。解决这一问题的方法可将明文的字母分组并进行编码，Hill 密码就是这样的一种系统。

设有明文信息

I　AM　HIDING。

要求 Hill 密码，先将明文分成对，并补哑字母 G 到最后，得

IA　MH　ID　IN　GG。

通常字母与数字对应表为已知，见表（*）。由表（*）得

91　138　94　914　77。

若采用矩阵 $\begin{bmatrix}1 & 2\\0 & 3\end{bmatrix}$ 编码，在编码时，取模为 26 的同余，则有

$$\begin{bmatrix}1 & 2\\0 & 3\end{bmatrix}\begin{bmatrix}9\\1\end{bmatrix}=\begin{bmatrix}11\\3\end{bmatrix}\to \text{密文为 KC}$$

$$\begin{bmatrix}1 & 2\\0 & 3\end{bmatrix}\begin{bmatrix}13\\8\end{bmatrix}=\begin{bmatrix}29\\24\end{bmatrix}\equiv\begin{bmatrix}3\\24\end{bmatrix}(\bmod\ 26)\to \text{密文为}$$

CX

$$\begin{bmatrix}1 & 2\\0 & 3\end{bmatrix}\begin{bmatrix}9\\4\end{bmatrix}=\begin{bmatrix}17\\12\end{bmatrix}\to \text{密文为 QL}$$

$$\begin{bmatrix}1 & 2\\0 & 3\end{bmatrix}\begin{bmatrix}9\\14\end{bmatrix}=\begin{bmatrix}37\\42\end{bmatrix}\equiv\begin{bmatrix}11\\16\end{bmatrix}(\bmod\ 26)\to \text{密文为}$$

KP

$$\begin{bmatrix}1 & 2\\0 & 3\end{bmatrix}\begin{bmatrix}7\\7\end{bmatrix}=\begin{bmatrix}21\\21\end{bmatrix}\to \text{密文为 UU}$$

通常将以不留间隔的单一字串来传递，按字母与数字对应表(*)查得这字串为：

KCCXQLKPUU。

字母与数字对应表(*)

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
P	Q	R	S	T	U	V	W	X	Y	Z				
16	17	18	19	20	21	22	23	24	25	0				

3.4 变化率思想与边际分析、弹性分析

1. 变化率的思想方法

函数在一点处的变化率又称函数在这点处的导数。它是指函数在这点处的改变量与自变量的改变量之比的极限。其思维实质是舍弃了事物具体运动形式的差别，仅考虑一变量相对于另一变量的变化快慢程度，从而具有很高的抽象程度。这种变化程度的刻画，借助了两增量之比与极限过程。由于增量之比反映的是不同事物的作用具有异质性，加之极限过程在本质上“不仅仅表明状态，并且也表明过程：运动”，故这种刻画已达到了十分精细和深刻的程度。

变化率的概念是一种思维原则，它引导人们通过考察两个相关变量的改变量之比的极限去分析事物的运动，且这种考察并非只借助定义，而主要是借助一套由定义建立起来的求导公式来完成，致使这样的考察具有很

高的思维效能。同时，由于函数关系的普遍性和广泛性，使得变化率在赋予每一具体函数关系时，就获得了种种具体的意义，从而具有广泛的使用价值。例如，当给出路程与时间的函数关系时，变化率就是变速运动的速度，而当给出平壁传热中温度与距离的关系时，变化率就是温度梯度等。导数本身，作为思维工具，它鲜明地刻画了函数在增减、凹凸、极值等方面的重要特性，而被人们广泛地应用着。

例如，以每分钟 4 立方米的速度将水注入深、上顶直径均为 8 米的倒圆锥形容器，求水深 5 米时，水面上升的速度(图 3.4)。

易知，水面上升的速度是水面深度 h 的函数，则函数关系是

$$V=\frac{1}{3}\pi\left(\frac{h}{2}\right)^2\cdot h=\frac{\pi}{12}h^3$$

要求深度 h 对时间 t 的变化率 h_t'，由于

$$V_t'=\frac{\pi}{12}\cdot 3h^2\cdot h_t',$$

故 $h_t'=\dfrac{4V_t'}{\pi h^2}$。

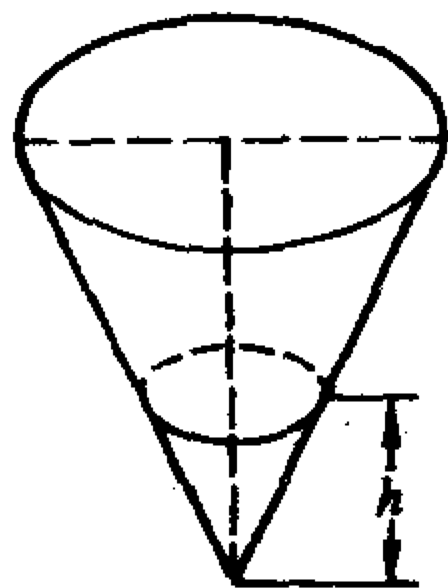


图 3.4

将 $Vt'=4, h=5$ 代入上式便得

$$h_t'=\frac{16}{25\pi}(\text{米}/\text{分})。$$

2. 边际分析 —— 变化率在经济中的应用(一)

在经济学研究中，若给出的函数分别代表成本、收入、利润等，则对应的导数就是所给出函数的边际函数。要对经济与企业的经营管理进行数量分析，“边际”是个重要概念，它是变化率在经济理论中的化身。

(1) 边际成本、平均成本。

某产品的总成本是指生产一定数量的产品所需的全部经济资源投入，包括劳力、原料、设备等的价格或费用总额。它由固定成本与可变成本组成，通常它是产量的函数。所谓边际成本就是总成本对产量的变化率。而平均成本乃是指生产一定量产品，平均每单位产品的成本。设 Q 为产量，$C(Q)$ 为总成本，则平均成本是 $\frac{C(Q)}{Q}$，边际成本为 $C'(Q)$，而 $C'(Q_0)$ 是 $C(Q)$ 在 Q_0 时的边际成本，它表示当产量为 Q_0 时，增加一个单位产品时所增加的总成本。

典型的成本函数是三次多项式。例如，

$$C = C(Q) = 0.001Q^3 - 0.3Q^2 + 40Q + 1000,$$

它的平均成本函数为

$$\overline{C}(Q) = 0.001Q^2 - 0.3Q + 40 + \frac{1000}{Q}。$$

边际成本函数为

$$C'(Q) = 0.003Q^2 - 0.6Q + 40。$$

当 $Q = 50$、100 时，$C'(50) = 17.5$，$C'(100) = 10$，分别给出了当生产水平为 50、100 时，总成本的瞬时变化率。

该边际成本函数 $C'(Q)$ 的图象是一条二次抛物线（图 3.5）。顶点为 $(100,10)$，在 $(0,100)$ 内单调下降，$(100,+\infty)$ 内单调上升。其下降部分的经济意义是：在生产很小的情况下，增加生产时增产部分的成本将更低，这是因为大批生产可以使生产能力得到充分利用，因此比少量生产更加经济，但对生产能力已定的工厂，大批生产的经济性不是无

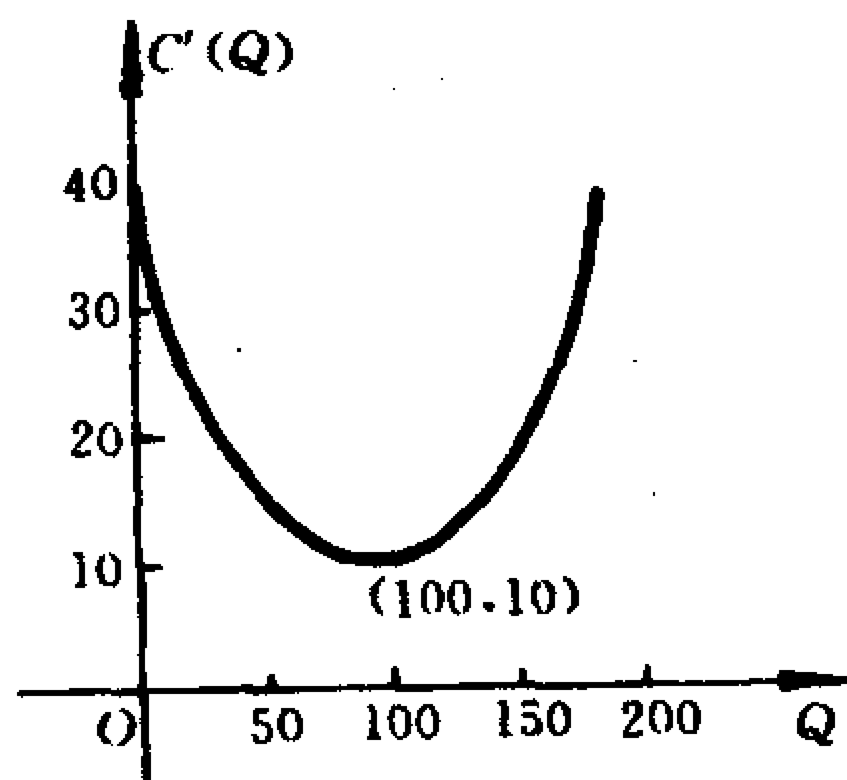


图 3.5

限的，当产量大到某一确定点后(如本例中 $Q=100$)，边际成本反会增加，这是因为在生产能力充分利用后，再要增加生产就必须投资新的机器、设备和增加人员，或者使机器和人员超负荷运转。这些都将导致成本的迅速上升，这在平均成本函数 $\overline{C}(Q)$ 中是难以看到的。

但边际与平均的概念也有联系。就总成本函数 $C=C(Q)$ 而言，对平均成本函数 $A_C=\dfrac{C(Q)}{Q}$ 对产量 Q 求变化率为：

$$\frac{d}{dQ}\left(\frac{C(Q)}{Q}\right)=\frac{1}{Q}\left(C'(Q)-\frac{C(Q)}{Q}\right),$$

故

$$\frac{d}{dQ}\left(\frac{C(Q)}{Q}\right)\begin{cases}>0,\text{当 } C'(Q)>\dfrac{C(Q)}{Q}\text{ 时},\\[2mm] =0,\text{当 } C'(Q)=\dfrac{C(Q)}{Q}\text{ 时},\\[2mm] <0,\text{当 } C'(Q)<\dfrac{C(Q)}{Q}\text{ 时}。\end{cases}$$

其经济意义是：平均成本曲线 A_C 上某点 Q 处的斜率在边际成本曲线 M_C 与平均成本曲线 A_C 交点 Q^* 处为零值，而当 M_C 在 A_C 上方时的 Q 点处取正值，下方时取负值(图 3.6 作的是 $C(Q)=Q^3-12Q^2+60Q$ 的 $C'(Q)$ 和 $\overline{C}(Q)$)。

(2) 边际利润和最大利润原则。

所谓总利润是指总收益减去总成本所得的差，而总收益乃是生产者出售一定产量产品所得的全部收入。边际收益和边际利润分别是总收益和总利润的变化率。若记总收益、总利润函数分别为 $R=R(Q)$，$L=L(Q)$，则

$$L(Q)=R(Q)-C(Q),L'(Q)=R'(Q)-C'(Q)。$$

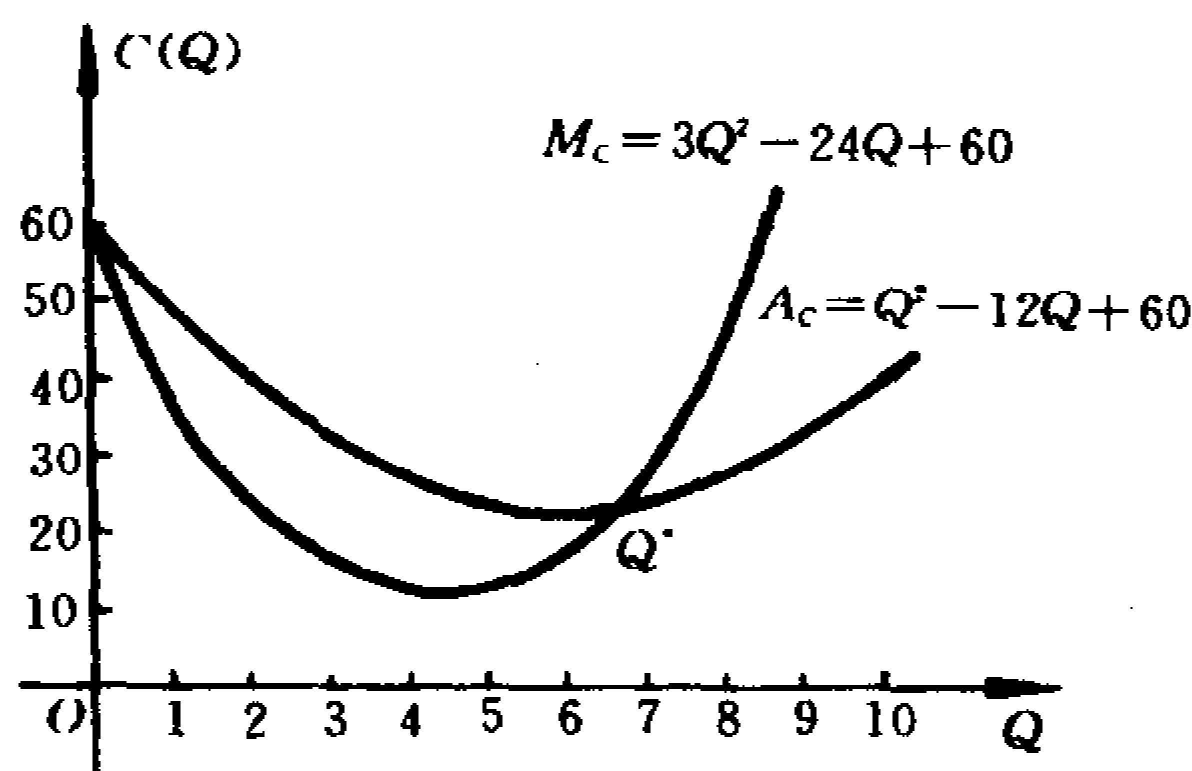


图 3.6

据此知 $L(Q)$ 取得最大值的必要条件是 $R'(Q)=C'(Q)$，亦即总利润 $L(Q)$ 最大的必要条件是边际收益等于边际成本。当 $L''(Q)<0$ 或 $R''(Q)<C''(Q)$，亦即边际收益的变化率小于边际成本的变化率时，总利润 $L(Q)$ 有最大值，这就是所谓的最大利润原则。

例如，某工厂生产某种产品，固定成本 20000 元，每生产一单位产品，成本增加 100 元，已知总收益 R 是年产量 Q 的函数：

$$R=R(Q)=\begin{cases}400Q-\dfrac{1}{2}Q^2, & 0\leqslant Q\leqslant 400,\\ 80000, & Q>400。\end{cases}$$

问每年生产多少产品时，总利润最大，并求其最大总利润。据题意，

$C=C(Q)=20000+100Q$，则总利润函数

$$L=L(Q)=R(Q)-C(Q)$$

$$=\begin{cases}300Q-\dfrac{1}{2}Q^2-20000, & 0\leqslant Q\leqslant 400,\\ 60000-100Q, & Q>400。\end{cases}$$

$$L'(Q)=\begin{cases}300-Q, & 0\leqslant Q\leqslant 400,\\ -100, & Q>400。\end{cases}$$

令 $L'(Q)=0$，得 $Q=300$。此时，$L''(Q)=-1<0$，所以 $Q=300$ 时 L 最大，且 $L(300)=25000$。即当年产量为 300 个单位时，总利润最大。其最大总利润为 25000 元。

3. 弹性分析 —— 变化率在经济中应用(二)

在经济学中，把某变量对另一变量变化反应的强烈程度或灵敏程度称为弹性。

(1) 弹性。

设有函数关系 $y=f(x)$，它在点 $x=x_0$ 处可导，函数的相对改变量 $\frac{\Delta y}{y_0}$ 与自变量的相对改变量 $\frac{\Delta x}{x_0}$ 之比 $\frac{\Delta y/y_0}{\Delta x/x_0}$，称为函数 $f(x)$ 从 $x=x_0$ 到 $x=x_0+\Delta x$ 两点间的相对变化率或称为弹性。当 $\Delta x\to 0$ 时，$\frac{\Delta y/y_0}{\Delta x/x_0}$ 的极限称为 $f(x)$ 在 $x=x_0$ 处的相对变化率或称为弹性。记作 $\left.\frac{Ey}{Ex}\right|_{x=x_0}$ 或 $\frac{E}{Ex}f(x_0)$。由函数在一点处弹性的定义知

$$\frac{Ey}{Ex}=\lim_{\Delta x\to 0}\frac{\Delta y/y}{\Delta x/x}=\lim_{\Delta x\to 0}\frac{\Delta y}{\Delta x}\cdot\frac{x}{y}=y'\cdot\frac{x}{y}。$$

它是 x 的函数，称为 $f(x)$ 的弹性函数。

根据上述定义，可求任一函数的弹性函数。

例如，$y=x^{\alpha}$（α 为常数），则

$$\frac{Ey}{Ex}=y'\frac{x}{y}=(\alpha x^{\alpha-1})\cdot\frac{x}{x^{\alpha}}=\alpha,$$

即是说，幂函数在任一点处的弹性都为固定的常数，故称幂函数为不变弹性函数。

(2) 需求弹性和供给弹性。

需求是指在一定价格条件下，消费者愿意购买并且有支付能力购买的商品量。影响商品需求的因素有多种，

除商品价格这一主要因素外，还有诸如消费者收入的增减，代用品等。现忽略其他因素，只考虑它与商品价格的关系。设 P 表示商品价格，Q 为需求量，则需求函数为 $Q=f(P)$，需求弹性函数为

$$\eta(P)=\frac{EQ}{EP}=\frac{f'(P)P}{f(P)}。$$

由于价格上涨时，需求一般总是下跌的，故 $\eta(P)$ 为负值。

例如，设某商品需求函数为 $Q=e^{-\frac{P}{5}}$，则需求弹性函数为

$$\eta(P)=\frac{EQ}{EP}=\frac{f'(P)\cdot P}{f(P)}=\frac{\left(-\frac{1}{5}e^{-\frac{P}{5}}\right)P}{e^{-P/5}}$$

$$=-\frac{1}{5}P。$$

而

$$\eta(3)=-\frac{3}{5}=-0.6,\eta(5)=-\frac{5}{5}=-1,$$

$$\eta(6)=-\frac{6}{5}=-1.2。$$

其经济意义是：

$\eta(5)=-1$，说明当 $P=5$ 时，价格与需求变动的幅度相同，即价格上涨 1%，需求下跌 1%；反之价格下跌 1%，需求上涨 1%。

$\eta(3)=-0.6$，说明当 $P=3$ 时，需求变动幅度小于价格变动幅度，即价格上涨 1%，需求下跌 0.6%。

$\eta(6)=-1.2$，说明当 $P=6$ 时，需求变动幅度大于价格变动幅度，即价格上涨 1%，需求下跌 1.2%。

经济学中，当 $\eta<-1$ 时，称需求是相当有弹性的，即价格的变化将引起需求的较大变化；当 $-1<\eta<0$ 时，

称需求是相当无弹性的，这时价格的变化只引起需求的微小变化，其需求量主要不是由价格决定；当$\eta=-1$时，称需求是有单位弹性的，即价格上涨与需求下降的百分数相同。由此可知，需求弹性的绝对值越大，需求对价格的依赖也越大，反之则越小。故需求弹性在价格商品理论中具有十分重要的意义。

与需求函数相对应的有供给函数。所谓供给是指在一定价格条件下，生产者愿意出售并有可供出售的商品量。设P表示商品价格，Q表示供给量，则供给函数为$Q=\varphi(P)$。供给弹性为

$$\varepsilon(P)=\varphi'(P)\frac{P}{\varphi(P)}。$$

由于供给函数是单调上升的，故供给弹性为正值。当$\varepsilon(P)>1$时，称供给是相当有弹性的；当$0<\varepsilon(P)<1$时，称供给是相当无弹性的；而当$\varepsilon(P)=1$时，称供给是有单位弹性的。

(3) 用弹性需求分析总收益的变化。

总收益R是商品价格P与销售量Q的乘积。即

$$R=P\cdot Q=P\cdot f(P),$$

$$R'=f(P)+pf'(P)=f(P)\left[1+f'(P)\frac{P}{f(P)}\right]$$
$$=f(P)(1+\eta)。$$

其经济意义是：

当$-1<\eta<0$时，需求变动幅度小于价格变动幅度，此时，$R'>0$，R递增，亦即价格上涨，总收益增加，价格下跌，总收益减少。

当$\eta<-1$时，需求变动幅度大于价格变动幅度，此时$R'<0$，R递减，即价格上涨，总收益减小，价格下跌，总收益增加。

当 $\eta=-1$ 时，需求变动幅度等于价格变动幅度，此时，$R'=0$，R 取得最大值。

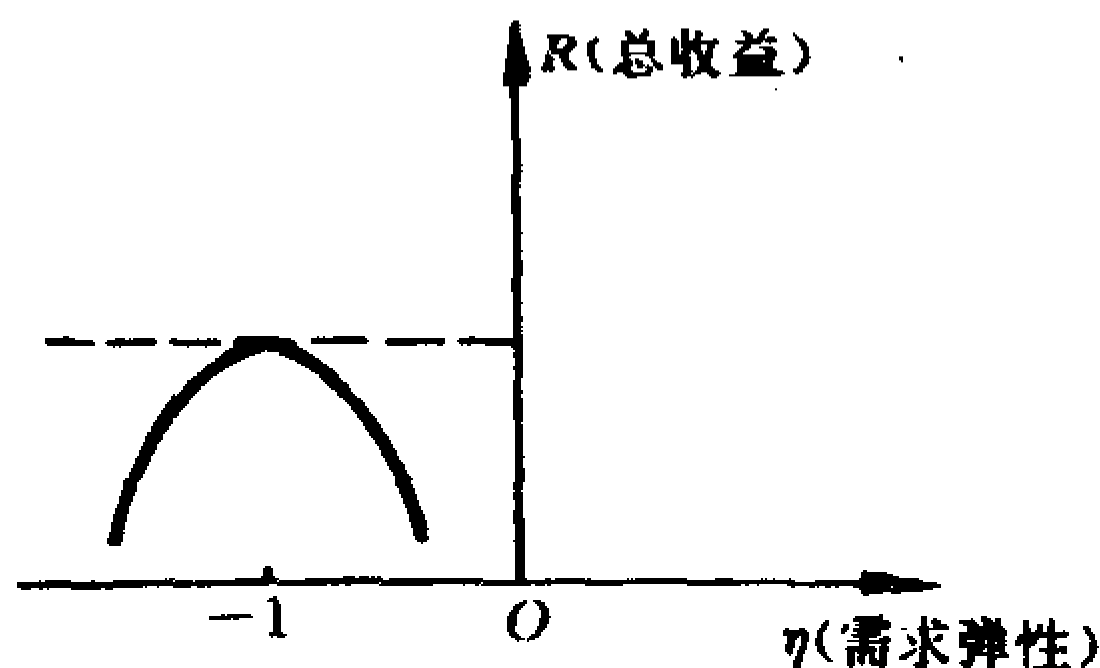


图 3.7

例如，设某商品需求函数为

$$Q=f(P)=12-\frac{P}{2}。$$

$$\eta(P)=f'(P)\frac{P}{f(P)}=\left(-\frac{1}{2}\right)\cdot\frac{P}{12-\frac{P}{2}}$$

$$=\frac{P}{P-24},$$

$$R=P\cdot f(P)=P\left(12-\frac{P}{2}\right),R'=12-P。$$

或直接代公式，得

$$R'=f(P)(1+\eta)=\left(12-\frac{P}{2}\right)\cdot\left(1+\frac{P}{P-24}\right)$$

$$=12-P。$$

令　$R'=0$，得 $P=12$，$\eta(12)=-1$，

$$R(12)=12(12-6)=72。$$

此时，因 $R''=-1<0$，故总收益有最大值，其最大总收益为 72，亦即在 $\eta(12)=-1$ 时，总收益 R 有最大值 72。

3.5 试验设计思想与正交试验方法

1. 试验设计的思想方法

试验设计的基本目的是为试验设计出正确有效的实施途径和实施方法，使得能由很少的试验次数就能获得良好的试验结果。安排试验的原则最为基本的是：均匀、对称、整齐、可比。所谓均匀、对称，是指试验的布点要有代表性，因素和水平间的搭配要合理，且试验要能反映出全局和整体的情况；所谓整齐、可比是指试验的优劣要能进行定量的分析和比较。

例如，正交试验设计是按正交表来安排试验的，而这种正交表是在正交拉丁方的基础上构造出来的，它具有如下特点：

1° 每列各字母出现的次数相同；

2° 任何两行横向组成各种有序数对出现的次数也相同。

如，正交表 $L_9(3^4)$ 有 9 行 4 列，每列字母 1、2、3 均出现 3 次，且任何两列横向组成的 9 个有序数对 (1,1)，(1,2)，(1,3)，(2,1)，(2,2)，(2,3)，(3,1)，(3,2)，(3,3) 各出现一次。故对于任何两列，字母“1”、“2”、“3” 间搭配是均匀的。又如，正交表 $L_8(4\times 2^4)$ 有 8 行 5 列，第一列由数字“1”，“2”、“3”、“4” 组成，每个数字各出现两次，其余各列都由数字“1”、“2” 组成，它们各出现 4 次，即各列本身每个数字出现的次数相同，且第一列与其余各列横向组成的 8 个有序数对 (1,1)、(1,2)、(2,1)、(2,2)、(3,1)、(3,2)、(4,1)、(4,2) 各出现一次，其余各列任何两列横向组成的有序数对 (1,1)、(1,2)、(2,1)、(2,2) 各出现两

次。故任意两列各种数字的搭配是均匀的。

这样的性质就是所谓的正交性。它保证了每个因素的各个水平参加试验的次数一样多，并且横向间搭配均匀。同时使所作的少数试验能较好地代表所有组合的情况，因此它们是均匀分布的。由于正交表的性质决定了正交试验具有整齐可比性，即它使每个因素在各个水平的结果之和中，其他因素的影响相同，从而可以分析、比较，以确定出主要、次要因素，好水平、不好水平，且作出优化方案来。正交表的这种均匀分散、整齐可比性，使正交试验实现了试验次数的以少代多。如正交表 $L_9(3^4)$，原需 $3^4 = 81$ 次试验，现只需做 9 次。

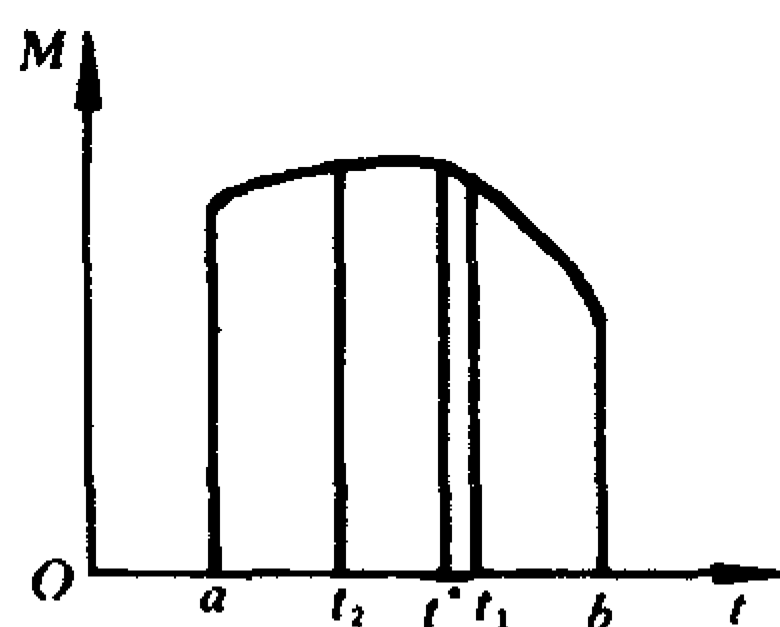
在黄金分割选优法中同样体现了这一思想。设要求函数 $M = f(t)$ 在区间 (a,b) 的最大(或最小值)。但函数 $M = f(t)$ 没有具体表达式，故对于每个 t 值所对应的 $f(t)$ 都只能通过试验求得，这是优选法的特点之一。令在 (a,b) 内任取两点：$a < t_2 < t_1 < b$，分别在 t_1、t_2 处做试验，求得对应的 $M_1 = f(t_1)$ 和 $M_2 = f(t_2)$，比较 M_1 和 M_2。若设 $M = f(t)$ 在 (a,b) 上只有一个最优点 t^*，又若 t_1 和 t_2 在 t^* 的两侧(图 3.8(1))，则试验后不论去掉 (a,t_2) 还是 (t_1,b)，t^* 都在留下部分内。若在 t^* 同侧，例如在右侧(图 3.8(2))，则试验结果必然 t_2 比 t_1 好，故丢掉 (t_1,b) 后，t^* 仍不会丢掉。因此，沿着“坏”点丢掉不包括好点在内的一段，最优点总不会丢掉，但试验范围却缩小了。同时，由于在未试验前，二次试验后哪个好是未知的，故在整个试验范围内的二次试验，丢掉 (a,t_2) 或 (t_1,b) 机会是相等的。这就要求它们等长，从而 $t_2 - a = b - t_1$，即 $t_2 = a + b - t_1$。同时为了保证试验从总体上是均衡、匀称的，每次留下区间的长度都是上次留下长度的 $w(0 < w <$

1）倍。设 t_1、t_2 两次试验后丢掉 (t_1,b) 留下 (a,t_1)，再在留下的区间 (a,t_1) 中取 $t_3 < t_2$，在 t_3 做试验得 $M_3 = f(t_3)$，比较后若 t_3 优，去掉 (t_2,t_1)，留下 (a,t_2)。若记 l_n 为第 n 次比较后留下的区间长度，则应有

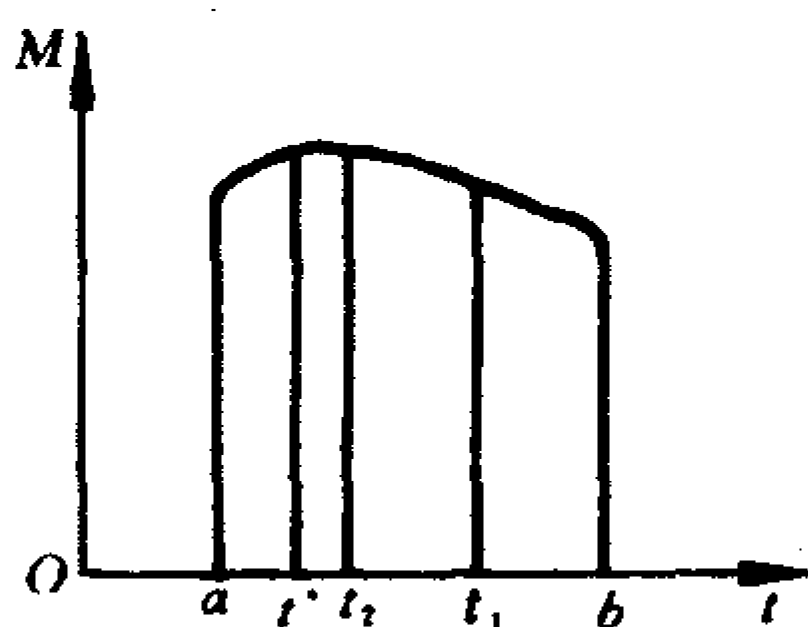
$$l_{n+1} = w \cdot l_n。$$

注意到 $l_0 = b - a, l_1 = t_1 - a$，又 $t_2 - a = b - t_1$，故

$$\begin{aligned} l_0 - l_1 &= (b - a) - (t_1 - a) = b - t_1 \\ &= t_2 - a = l_2。\end{aligned}$$



(1)



(2)

图 3.8

但 $l_1 = wl_0, l_2 = wl_1$，故

$$l_0 - wl_0 = w(wl_0),$$

因 $l_0 \neq 0$，可得

$$w^2 + w - 1 = 0。$$

由于 $0 < w < 1$，则

$$w = \frac{-1 + \sqrt{5}}{2} \approx 0.6180339,$$

它就是黄金分割的比值，通常称为黄金数。

用黄金分割比值安排优选试验可以以较少的试验次数获得理想的试验效果。理由在于这种安排体现了试验点的均衡、匀称、和谐、协调。美学研究表明，和谐、匀称、

比例的合度具有美感。由于黄金分割使整体和部分达到了和谐，匀称和比例的合度，故用黄金数分割对象，可使对象具有美感。曾有人作过试验，让人们在不同的长方形中挑选一个美的图形，结果绝大部分人选中的书籍、广告等的宽、长比取为黄金数。在建筑中，在高塔旁建造楼阁和设置平台，总把楼阁和平台设置在高塔的黄金分割点处，因为这样设计可使瘦削单调的塔身变得丰富多彩。同样，在摩天大楼的黄金分割点处布置腰线或装饰物，整个楼房就显得宏伟、雅致。在摄影和美术创作中，一般不把表现的主题放在画面的正中，而应放在偏离中心的黄金分割点处。在雕塑中，以人的下肢与身高之比取为黄金分割比时为最美。古希腊的智慧女神雅典娜和太阳神阿波罗的塑像，都是采用这种身段比。舞台上，歌唱演员若站在舞台正中就会显得刻板，而站在黄金分割点处，则显得活泼、和谐，且音响效果也较好。

历史上，对黄金分割的研究由来已久。相传这一分割是由古希腊人首先发现的，后来，不少人又相继作了研究。15 世纪，意大利画家波堤切利著有《论神圣的比例》，认为数学中的这个比例是地球上的一切称为美的东西都必须遵循的。17 世纪著名科学家开普勒将“黄金分割”和“勾股定理”誉为几何学的两个罕有的宝藏。文艺复兴时期的美术大师达·芬奇，不仅在自己的艺术实践中广泛地应用了这个分割，而且用数学方法探讨了应用解剖、透视、构图等方面的系统理论，并把它命名为黄金分割。自此，这一名称便一直延用至今。

2. 正交试验法

正交试验法是指应用正交表进行试验的设计和对试验的结果进行数量分析的一种试验方法。这种方法的思

路是从实际问题中确定出试验的因素和水平 —— 选择恰当的正交表，并将因素和水平安排在正交表上，得出试验方案 —— 按试验方案试验，并记录试验数据 —— 分析试验结果，以确定出推广的最佳方案。

例如，某化工厂为提高产品回收率，用正交试验法寻找合理的生产工艺。据试验和分析知，影响回收率的原因中，反应温度、加碱量和催化剂种类三种最为重要。故挑选这三种因素进行试验。对每个因素，确定其相应的试验范围，定出若干水平，得如下的因素水平表：

水平＼因素	A 反应温度(℃)		B 加碱量(kg)		C 催化剂种类	
1	A_1	80	B_1	35	C_1	甲
2	A_2	85	B_2	48	C_2	乙
3	A_3	90	B_3	55	C_3	丙

选择 $L_9(3^4)$ 表，将 A、B、C 三因素分别置于表的1、2、3列，便得试验方案如下：

试验号＼因素	A(℃)		B(kg)		C		指标 收率(%)
1	1	(80)	1	(35)	1	(甲)	51
2	1		2	(48)	2	(乙)	71
3	1		3	(55)	3	(丙)	58
4	2	(85)	1		2		82
5	2		2		3		69
6	2		3		1		59
7	3	(90)	1		3		77
8	3		2		1		85
9	3		3		2		84
K_1	180		210		195		
K_2	210		225		237		

试验号＼因素	A(℃)	B(kg)	C	指标 收率(%)
K_3	246	201	204	
k_1	60	70	65	
k_2	70	75	79	
k_3	82	67	68	
极差 R	22	8	14	

通过试验，将试验的数据填入表中右边的指标栏中，并对试验结果进行分析，分别求出各因素的对应水平试验结果的和。例如，第一列中，温度为 80℃ 的三种指标的和为 $K_1 = 51 + 71 + 58 = 180$，温度为 85℃、90℃ 的三种指标的和分别为 $K_2 = 210$，$K_3 = 246$，将它们记入该列表的 $K_i(i = 1、2、3)$ 栏中。同理求出 B、C 对应的 K_i。表的 k_i 是每个因素各水平条件下试验的平均指标，可从 K_i 求得。又称各列表中 k_i 的最大、最小值之差为极差，记为 R。极差的大小反映了同一因素取不同水平时，试验指标的变化幅度。极差越大，表明因素对试验指标的影响也越大，因而就越重要。这里，A 的极差(22) 最大，C 为次，B 最小。故三个因素主次顺序为 $A—C—B$。在各列的考核指标中，A_3、C_2、B_2 分别最好，故它可能是 A、B、C 的最佳试验条件，而这个试验不包括在已做的 9 个试验中，因而要作验证试验。取 9 个试验中最高收率的 8 号试验与其作对比。验证试验结果知 A_3、B_2、C_2 的收率 92.5% 高于 8 号的试验收率，于是知 A_3、B_2、C_2 为最佳试验条件。同时，从上述试验中看出，温度越高，收率也越高，在可能条件下，如将反应温度再适当提高作试验，也还可能有较高的收率。

正交试验法是处理各因素选优问题的一种重要设计

和分析方法，它可以以较少的试验次数来获得较好的试验结果。其原理在于如前面所说的，试验点分布的均匀分散性。在本例中，原做 27 次的全面试验（试验点在立方体表面的 27 个点上），改用正交试验法，只做了 A_1、B_1、C_1；A_1、B_2、C_2；A_1、B_3、C_3；A_2、B_1、C_2；A_2、B_2、C_3；A_2、B_3、C_1；A_3、B_1、C_3；A_3、B_2、C_1；A_3、B_3、C_2 等 9 次，这 9 个点均匀分布在立方体的三个面上，从而能全面地反映出整个立方体的情况（图 3.9）。

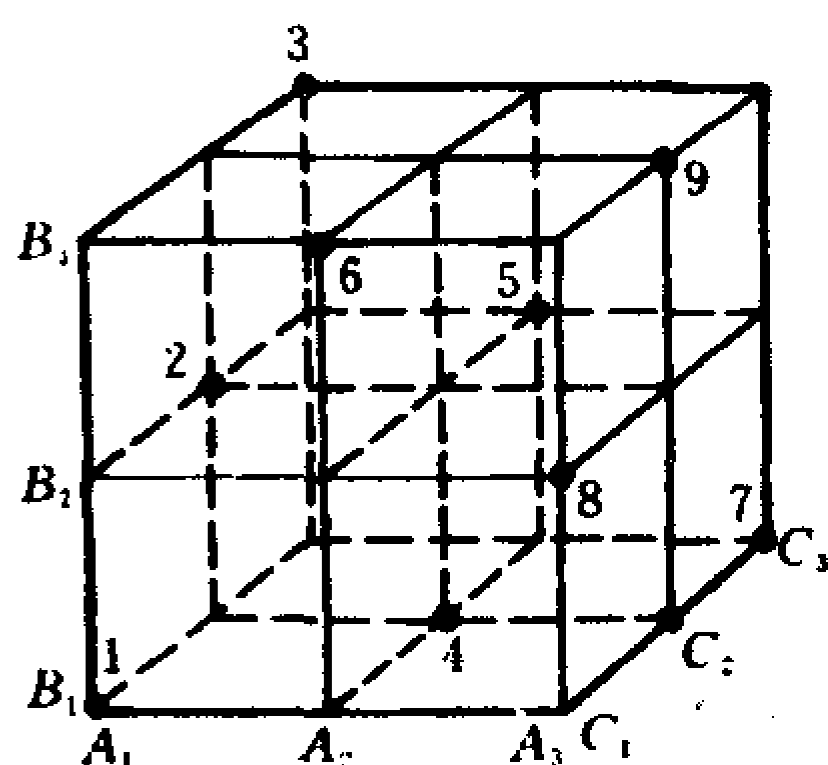


图 3.9

目前，正交试验法已大量应用于工农业生产。如据《数理统计与管理》卷 12 第 1 期刊文，河北工学院郝万德、刘金海针对热镀锌生产中灰铸铁锌锅易腐蚀，应用正交试验法进行熔铸渗硼试验，找到了较好的工艺条件，即在这样的工艺条件下熔铸渗硼，使灰铸铁件表面形成渗硼层，解决了灰铸铁锌锅的腐蚀问题，使灰铸铁锌锅的使用寿命提高了 25 倍。据郝、刘调查，河南省某些工厂的灰铁锅原每月要换 7 口，一口这样的锅价值 2000 元。故使用正交试验找到的新工艺，对降低热镀锌生产工艺的成本作用是很大的。

3.6 群结构原理与几何学

1. 群的概念和原理

所谓群是指这样的一个集合 G，如果在这集合中可引进运算“$\circ$”，且满足以下四条性质：

1° 封闭性：对于 $\forall a、b \in G$，有 $a \circ b \in G$。

2° 结合性：即对于 $\forall a、b、c \in G$，有

$$(a \circ b) \circ c = a \circ (b \circ c) = a \circ b \circ c。$$

3° 存在单位元素 $e \in G$，即对于 $\forall a \in G$，有

$$a \circ e = e \circ a = a。$$

4° 存在逆元素，即对于 $\forall a \in G$，存在元素 $a^{-1} \in G$，使得

$$a \circ a^{-1} = a^{-1} \circ a = e。$$

则称集合 G 关于运算 $\circ$ 是一个群。如果这个群的运算还满足

5° 交换律：即对于 $\forall a、b \in G, a \circ b = b \circ a$，

则这个群称为交换群或 Abel 群。

例如，整数 Z 对普通加法运算构成一个交换群，因为它满足运算的封闭性、结合性且有单位元 0；又整数 a 的逆元是其相反数 $-a$。同理，正有理数集 Q^+ 对于普通乘法也构成一个 Abel 群。单位元素是正整数 1，任一正有理数 a 的逆元是它的倒数 $\frac{1}{a}$。但全体负有理数集 Q^- 对于乘法就不构成群，因为它对乘法运算不封闭。又正整数对乘法运算也不构成群，因为任一正整数，其倒数不是正整数，故它没有逆元。

又如，在全体整数上定义模 2 加法 $\oplus$：即对于任意整数 $a、b$，如果

$$a + b \equiv 0(\mathrm{mod}\ 2),$$

即 $a+b$ 为偶数，则 $a \oplus b = 0$。又

$$a + b \equiv 1(\mathrm{mod} 2),$$

即 $a+b$ 为奇数，则 $a \oplus b = 1$。

由于整数按“模 2 相等”的概念只有两个元素 0 和 1，

故实际只有以下的情况：

$$0 \oplus 0 = 0, 0 \oplus 1 = 1, 1 \oplus 0 = 1, 1 \oplus 1 = 0。$$

因而全体整数关于模 2 的加法构成群，其单位元是 0，0 和 1 的逆元分别是自身。

再如，设有集合{1,2,3,4}，考虑 4 个字母的所有可能的置换所成的集合。即由

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \end{pmatrix} = I_0, \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 3 & 4 \end{pmatrix} = I_1, \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 2 & 1 & 4 \end{pmatrix} = I_2, \cdots,$$

$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 2 & 1 \end{pmatrix} = I_{23}$ 所组成的置换集合，该集合中有 24 个元素。可验证它满足封闭性、结合性，且有单位元 I_0。因为集合中的任一置换与 I_0 之积仍为这一置换。如

$$I_{23} \cdot I_0 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 2 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 2 & 1 \end{pmatrix} = I_{23}。$$

又，任一置换都有逆置换在这集合中。如 I_1 的逆置换为 I_1，因为

$$I_1^2 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 3 & 4 \end{pmatrix} \cdot \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 3 & 4 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \end{pmatrix} = I_0,$$

故$\{I_0, \cdots, I_{23}\}$ 构成一个置换群，这是一个 24 阶的置换群，记为 S_4。

群的概念不仅与集合有关，而且与所考察的运算有关。群概念揭示的是数学对象及其它们的相互关系，亦即对象的关系结构，这种关系结构已经抽去了对象的具体内容，而只考虑对给定意义下的运算，凡满足封闭性、结合性、有单位元、有逆元的任何一种抽象物的集合。它是建立结构理论的基础。瑞士著名心理学家皮亚杰正是应用了群概念的思想实质，建立了结构主义理论。他在其名著《结构主义》中阐述了这一理论的基本原理是：

1° 结构的整体性；

2° 结构的转换规律；

3° 结构的自身调整性。

皮亚杰在论述这些原则时，是从群的概念出发的。他认为，正是由于群公理所规定的性质，才保证了结构各部分的联系、转换机制和系统的相对封闭性等。

这种以群论原理建立起来的各种具体结构在实践中得到了广泛的应用。例如，相对论中的洛仑兹群，量子力学中的李群等都是现代科学中的常识性工具。数学中高次方程式根的求解，也与这一概念相紧密联系着。

例如，考察四次方程

$$x^4 + ax^3 + bx^2 + cx + d = 0。$$

由韦达定理的根与系数关系，得

$$\begin{cases} x_1 + x_2 + x_3 + x_4 = -a, \\ x_1x_2 + x_1x_3 + x_1x_4 + x_2x_3 + x_2x_4 + x_3x_4 = b, \\ x_1x_2x_3 + x_1x_2x_4 + x_2x_3x_4 + x_1x_3x_4 = -c, \\ x_1x_2x_3x_4 = d。 \end{cases}$$

以上四式左端为基本对称多项式，对置换群 S_4 不变。若作多项式 $V_1 = x_1 + x_2 - x_3 - x_4$，在 S_4 的作用下，V_1 的 24 种变化只有 6 种不同。例如，分别在

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \end{pmatrix}、\begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 3 & 4 \end{pmatrix}、\begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \end{pmatrix}、\begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 4 & 3 \end{pmatrix}$$

作用下，V_1 仍变为 V_1。又分别在

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 1 & 2 \end{pmatrix}、\begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 1 & 2 \end{pmatrix}、\begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 2 & 1 \end{pmatrix}、\begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 2 & 1 \end{pmatrix}$$

作用下，V_1 变为

$$V_2 = -x_1 - x_2 + x_3 + x_4 = -V_1。$$

同理，另外四种是

$V_3 = x_1 + x_3 - x_2 - x_4, V_4 = - x_1 - x_3 + x_2 + x_4 = - V_3, V_5 = x_1 + x_4 - x_2 - x_3, V_6 = - x_1 - x_4 + x_2 + x_3 = - V_5$。

作关于 S_4 不变的对称多项式构成的方程：

$$(t - V_1)(t - V_2)\cdots\cdots(t - V_6)$$
$$= (t - V_1)(t + V_1)(t - V_3)(t + V_3)(t - V_5)(t + V_5)$$
$$= (t^2 - V_1^2)(t^2 - V_3^2)(t^2 - V_5^2) = 0。$$

因为它是关于 t^2 的三次方程，三次方程已有公式解，可得关于 t^2 的三个解，因而得 t 的六个解。

从四元一次方程组

$$\begin{cases} V_1 = x_1 + x_2 - x_3 - x_4, \\ V_3 = x_1 + x_3 - x_2 - x_4, \\ V_5 = x_1 + x_4 - x_2 - x_3, \\ - a = x_1 + x_2 + x_3 + x_4 \end{cases}$$

中解得

$$\begin{cases} x_1 = \dfrac{1}{4}(V_1 + V_3 + V_5 - a), \\ x_2 = \dfrac{1}{4}(V_1 - V_3 - V_5 - a), \\ x_3 = \dfrac{1}{4}(- V_1 + V_3 - V_5 - a), \\ x_4 = \dfrac{1}{4}(- V_1 - V_3 + V_5 - a)。 \end{cases}$$

而 V_1、V_3、V_5 可用基本对称多项式表示，又基本对称多项式的值恰为方程的系数，因此四次方程的根式求解就完成了。上述求解过程中，置换群的概念起了重要的作用。

这样的求解思想是否能适用于五次或五次以上的方程，这是个十分自然的问题。无数次的尝试都失败了，究其原因发现，代换

$$V = x_1 + x_2 - x_3 - x_4$$

是一个技巧，有了V的这一表达式，四次方程求解才能化为t^2的三次方程求解。法国数学家拉格朗日首先意识到，对五次方程，能起这样作用的V一般并不存在，他还于1770年指出，根的置换理论是解代数方程的关键，并将使函数保持不变的根的置换集称作群，这就是群这一词的最初含义。1824年挪威数学家阿贝尔(Abel)证明了五次方程一般不可解，但当时无人理解。不久，年轻的法国数学家伽罗瓦(Galois)证明：一个方程如果可以用根式求解，相应的Galois群必须属于一定的类，即Galois群是可解群。而相应于一般的n次方程的Galois群，只有在$n = 1,2,3,4$时才是可解群。这样，伽罗瓦完全解决了五次及五次以上根式求解一般不可能的问题，并在这过程中创立了群论。这种群论的重要性和意义，著名数学家波莱尔的如下一段话是耐人寻味的。他说："这种群的所有重要性质，例如可解性，实际上是不依赖于被置换对象的性质的，从而产生了'抽象群'的概念以及某些非常重要的定理，可以应用于许多领域。但是，多年来这似乎只是纯粹的抽象数学。1910年左右，Princeton大学一位数学家和一位物理学家在讨论课程表的时候，那位物理学家说，他们无疑可以删掉群论，因为它决不会对物理有用的。没过二十年，出版了三本关于群论与量子力学的书，此后，群也就成为物理学中十分重要的东西了。"[5]

2. 变换群与几何学

F. 克莱因(F. Klein)的几何观以"爱尔兰根纲领"(Erlangcn programm)闻名于世，这是克莱因1872年在德国爱尔兰根大学所作的题为"近世几何学研究的比较评论"的报告中提出的。他在演讲时明确地宣布说："几何

学尽管本质上是一个整体，可是，由于最近期间所取得的飞速发展，却被分割成为许多几乎互不相干的分科，其中每一个分科几乎都是独立地继续发展着，于是，公开发表旨在建立几何学的这样一种内在联系的各种考虑，就显得更加必要了。"[10]具体地说，F. 克莱因提出了如下的关于几何的新观点：

设给定一个集合以及此集合的元素间的一个变换群，我们把此集合叫做空间，集合的元素叫做点，子集叫做图形，于是空间内图形对于此群的不变性质（包括不变量与不变性）的命题系统的研究就称为这空间的几何学，而空间的维数就称为几何学的维数，且称此群为它的基本群。有一个变换群就相应地有一种研究在这个群作用下不变性质理论的几何学。例如，欧氏几何研究的就是正交变换群之下不变性质的命题系统。相应地，抛物度量几何是相似变换下的不变性质的命题系统，而仿射几何、射影几何对应的变换群则是仿射变换群和射影变换群。

为什么能用变换群来统一各种几何学？这是由变换群定义中的公理所决定的。以欧氏几何为例，它所研究的图形 F 的性质是不同的点与其合同的一切图形所共有的。这种在不同的点与 F 合同的图形组成的一类，正是我们熟知的合同等价类。而等价关系是必须满足反射律、对称律、传递律等三律的，但由于正交变换构成正交变换群，按群公理的要求，作为正交变换群中任一元素的正交变换必具有下列性质：

1°恒等变换是正交变换；

2°正交变换的逆变换是正交变换；

3°两个正交变换的积是正交变换。

这正交变换的三种性质恰与合同等价关系的三律完全相

对应，从而将欧氏几何看作为是正交变换下的不变性质的命题系统就是十分自然的事了。

克莱因的几何观还指出了各种几何的内在联系。正因为一个变换群决定一种几何学，而变换群作为一种定义了运算的集合，它们的外延具有包含与被包含的关系。若集合 S 的变换群 $G_B \subset G_A$，则说 G_B 是 G_A 的子群。这样，G_B 决定的几何系统就是 G_A 的子几何系统。反之，G_A 的几何系统是 G_B 的原几何系统。由于任一变换群的不变性质对于它的子群总是成立的，而反之却未必，故原几何系统的性质，子几何系统一定成立。反之，子几何系统性质原几何未必成立。若用 K、A、S、M 分别表示射影群、仿射群、相似群、正交群，则 $K \supset A \supset S \supset M$。故仿射几何是射影几何的子几何，而抛物度量几何、欧氏几何则分别是仿射几何和抛物度量几何的子几何。各种几何的比较见下表。

名　称	射影几何	仿射几何	抛物几何	欧氏几何
相应的变换群	射影群	仿射群	抛物度量群	正交群
变换式	$\rho x'_i = \sum_{j=1}^{3} a_{ij}x_j$ $(i=1,2,3)$ $\rho\lvert a_{ij}\rvert \neq 0$	$\begin{cases} x' = a_{11}x \\ \quad + a_{12}y + a_{13} \\ y' = a_{21}x \\ \quad + a_{22}y + a_{23} \end{cases}$ $\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \neq 0$	$\begin{cases} x' = ax \\ \quad - \lambda by + c_1 \\ y' = bx \\ \quad + \lambda ay + c_2 \end{cases}$ $\lambda = \pm 1$	$\begin{cases} x' = x\cos\theta \\ \quad - \lambda y\sin\theta + a \\ y' = x\sin\theta \\ \quad + \lambda y\sin\theta + b \end{cases}$ $\lambda = \pm 1$
参数数目	8	6	4	3
研究对象	射影性质 射影不变量	仿射性质 仿射不变量 同左	相似性质 相似不变量 同左	度量性质 度量不变量 同左
基本不变性质	结合性	结合性 平行性	结合性 平行性 相似性	结合性 平行性 合同性

名　称	射影几何	仿射几何	抛物几何	欧氏几何
基本不变量	交比	单比	线段之比	距离
基本不变图形	—	无穷远直线	{圆点 I,J} $I(1,i,0)$ $J(1,-i,0)$	{圆点 I,J}

了解了各种几何间的相互关系，可使我们获得居高临下看待和处理各种几何问题的能力。

例如；两图形的位似被定义为图形 F 上任一点 P 与对应图形 F' 上的点 P'，对某一固定点 O 有 P'、P、O 在一直线上，且 $(P'PO)=k\neq 0$，其中 $(P'PO)=\dfrac{\overrightarrow{P'O}}{\overrightarrow{PO}}$。由于单比 $(P'PO)$ 是仿射不变量，P'、P、O 共线是结合性，故这一概念是属于仿射几何的。当然，它对于仿射几何的子几何抛物度量几何、欧氏几何等亦适用，但对射影几何不适用。

又如，三角形垂心，它在相似变换下仍保持其垂心原有的面貌，从而对其子几何欧氏几何亦成立。但在仿射变换、射影变换下，三角形的垂心不能再保持其垂心原有的特征。

F. 克莱因将群公理与等价关系联系起来思考，使原先被分割开来的几何学在群的观念下重新被统一起来了。他的这一工作在数学史上获得了很高的评价。例如，数学史家 M. 克莱因就曾指出："虽然 F. 克莱因的几何观点不能证明无所不包，但它确实能给大部分的几何提供一个系统的分类方法，并揭示很多可供研究的问题。他的几何'定义'指引了几何思想约有 50 年之久。"[17]

在处理具体几何问题时，几何学的变换群观点往往注重于寻找图形在变换中的不变的数量和关系，而这常

常可以使问题的求解变得异常简洁。

例 1　结论隐含在图形的旋转对称之中。

在正方形 $A_1A_2A_3A_4$ 内取点 P，由顶点 A_1 向 A_2P 引垂线 l_1，由 A_2 向 A_3P 引垂线 l_2，由 A_3 向 A_4P 引垂线 l_3，由 A_4 向 A_1P 引垂线 l_4。证明：所引的四条垂线（或延长线）相交于一点（图 3.10）。

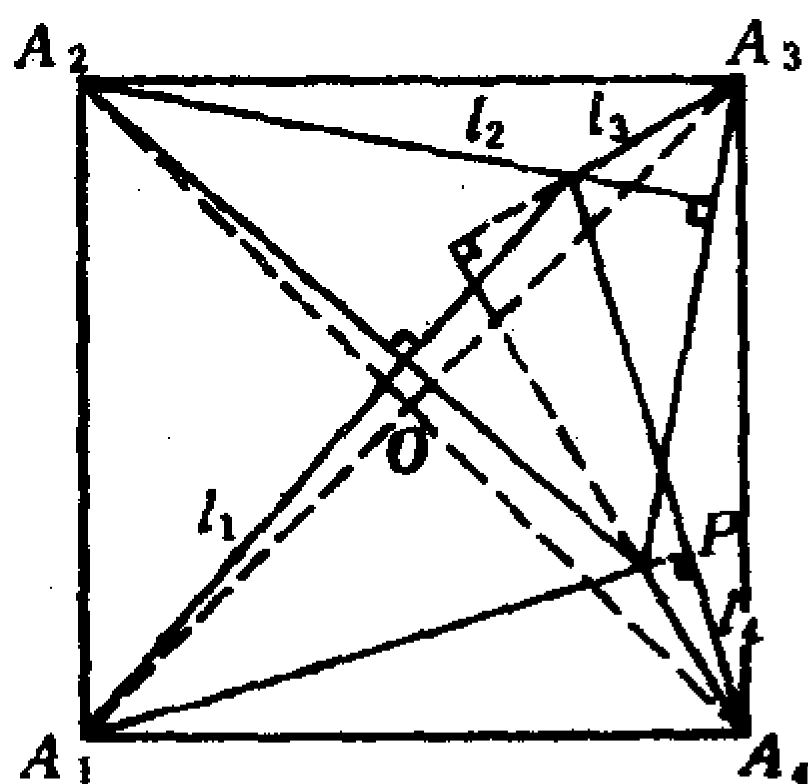


图 3.10

本题是四线共点问题，待证的四线 l_1、l_2、l_3、l_4 是另外四线 A_2P、A_3P、A_4P、A_1P 的垂线。鉴于所论图形是正方形，其对角线交点 O 是图形的旋转对称中心，若以 O 为中心，顺时针 90° 的旋转记为 $R(0,-90°)$，由于

$$A_1A_2 \xrightarrow{R(0,\,-90°)} A_2A_3,$$

故

$$l_1 \xrightarrow{R(0,\,-90°)} A_2P。$$

这是因为 l_1 经 $R(0,-90°)$ 后所得的直线必过 A_2 且垂直于 l_1，这样的直线只能是 A_2P。同理

$$l_2,l_3,l_4 \xrightarrow{R(0,\,-90°)} A_3P,A_4P,A_1P。$$

因为 $A_iP(i=1,2,3,4)$ 四直线共点于 P，又旋转不改变点与直线的结合关系，故 l_1、l_2、l_3、l_4 也必共点。

本题的求解抓住了旋转中的不变关系 —— 点线的结合关系和不变量 —— 直线的交角，解法简捷而明快。

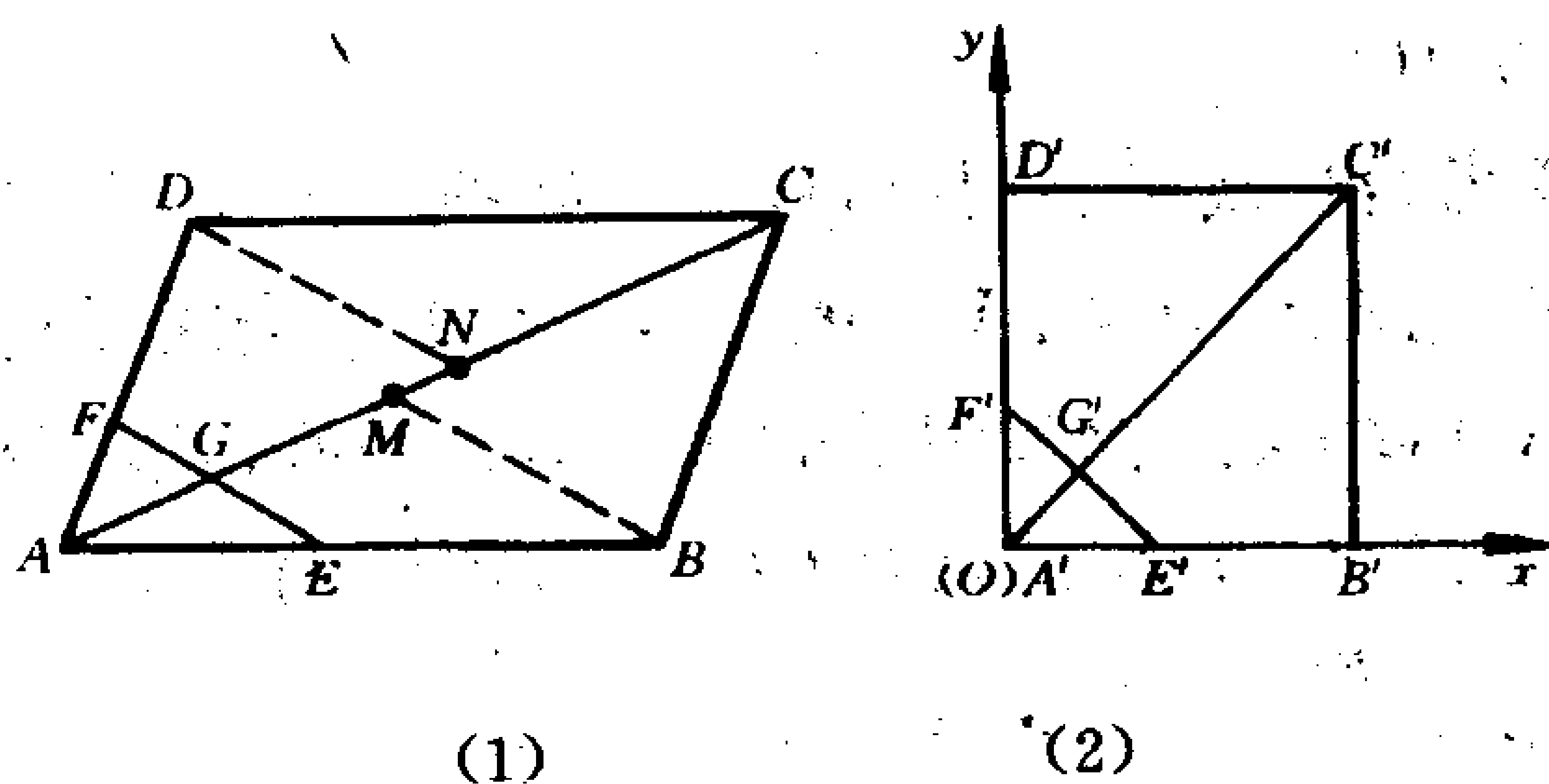
例 2　仿射成为特殊图形。

已知平行四边形 $ABCD$ 中，E、F 分别是 AB、AD 上

的点，且 $AE=\frac{1}{m}AB$，$AF=\frac{1}{n}AD$，EF 交 AC 于 G。若 $AG=\lambda AC$，求 λ 的值(图 3.11(1))。

问题是求一直线上两线段之比，这是仿射不变量。由于仿射变换可将任意三角形变为一特殊三角形，例如可变为等腰直角三角形。记 T 是将 $\triangle ABC$ 变为等腰直角三角形 $A'B'C'$ (B' 为直角顶) 的仿射变换。这时，

平行四边形 $ABCD \xrightarrow{T}$ 正方形 $A'B'C'D'$(图 3.11(2))，



(1) (2)

图 3.11

而

$$AG/AC=A'G'/A'C',$$

故要求 AG/AC，只需求 $A'G'/A'C'$。

若以 A' 为坐标原点，$A'B'$、$A'D'$ 分别为 x 轴、y 轴建立直角坐标系，取 $A'B'=A'D'=1$，则

$$A'C':y=x。$$

因为在仿射变换中，

$$AE/AB=A'E'/A'B',AF/AD=A'F'/A'D'。$$

故 $E'\left(\frac{1}{m},0\right)$，$F'\left(0,\frac{1}{n}\right)$，因而

$$E'F':\frac{x}{\frac{1}{m}}+\frac{y}{\frac{1}{n}}=1,\text{即 } mx+ny=1。$$

$$\begin{cases}y=x\\mx+ny=1\end{cases}\Leftrightarrow x=y=\frac{1}{m+n}。$$

故 $G'\left(\frac{1}{m+n},\frac{1}{m+n}\right)$。

因而

$$A'G'=\frac{1}{m+n}\sqrt{2}=\frac{1}{m+n}A'C'。$$

则 $AG=\frac{1}{m+n}AC$，即 $\lambda=\frac{1}{m+n}$。

本题的求解抓住了仿射变换的基本不变量——单比亦即仿射比。若不用仿射方法，亦可直接求解：

作 $BM\parallel EF$，$DN\parallel EF$。$\because\triangle ABM\cong\triangle CDN$，$\therefore AM=CN$。

又 $\frac{AG}{AM}=\frac{AE}{AB}=\frac{1}{m}$，$\frac{AG}{AN}=\frac{AF}{AD}=\frac{1}{n}$，故 $AM=m\cdot AG$，$AN=n\cdot AG$。

则 $AC=AN+NC=AN+AM=(n+m)AG$。

从而 $\lambda=\frac{AG}{AC}=\frac{1}{m+n}$。

直接求解虽然也很简洁，但要添置辅助线，技巧性较强，相比而言，仿射的方法既简洁，又较程式化，因而容易找到求解方法。

例 3 交比——射影几何的一种活力。

设 AB 为圆 O 直径，C 为 AB 延长线上一点，CM 切圆 O 于 M，$MH\perp AC$ 于 H，则 $\overline{AH}\cdot\overline{BC}=\overline{AC}\cdot\overline{HB}$。（$\overline{AH}$ 表示有向线段 AH）

连结 OM。在 Rt$\triangle OMC$ 中，$OM^2=OH\cdot OC$，则 $OA^2=OH\cdot OC$。

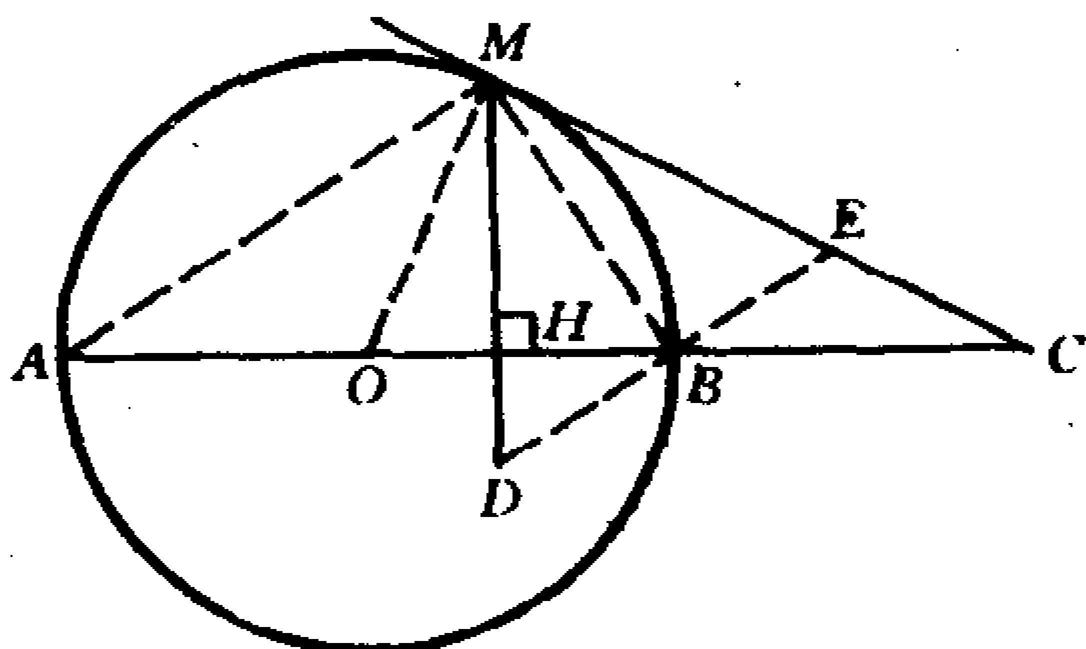


图 3.12

$$\because \quad \overrightarrow{OA} = -\overrightarrow{OB}, \quad \therefore \quad \overrightarrow{OA} + \overrightarrow{OB} = \overrightarrow{O},$$

故 $2(\overrightarrow{OA} \cdot \overrightarrow{OB} + \overrightarrow{OH} \cdot \overrightarrow{OC}) - (\overrightarrow{OH} + \overrightarrow{OC})(\overrightarrow{OA} + \overrightarrow{OB}) = 0$

$$\Rightarrow \overrightarrow{OA} \cdot \overrightarrow{OB} - \overrightarrow{OH} \cdot \overrightarrow{OB} - \overrightarrow{OA} \cdot \overrightarrow{OC} + \overrightarrow{OH} \cdot \overrightarrow{OC} + \overrightarrow{OA} \cdot \overrightarrow{OB} - \overrightarrow{OC} \cdot \overrightarrow{OB} - \overrightarrow{OA} \cdot \overrightarrow{OH} + \overrightarrow{OH} \cdot \overrightarrow{OC} = 0$$

$$\Rightarrow \overrightarrow{OB}(\overrightarrow{OA} - \overrightarrow{OH}) - \overrightarrow{OC}(\overrightarrow{OA} - \overrightarrow{OH}) + \overrightarrow{OB}(\overrightarrow{OA} - \overrightarrow{OC}) - \overrightarrow{OH}(\overrightarrow{OA} - \overrightarrow{OC}) = 0$$

$$\Rightarrow (\overrightarrow{OA} - \overrightarrow{OH})(\overrightarrow{OB} - \overrightarrow{OC}) + (\overrightarrow{OA} - \overrightarrow{OC})(\overrightarrow{OB} - \overrightarrow{OH}) = 0$$

$\Rightarrow \overrightarrow{HA} \cdot \overrightarrow{CB} + \overrightarrow{CA} \cdot \overrightarrow{HB} = 0$ 或 $\overrightarrow{AH} \cdot \overrightarrow{BC} = \overrightarrow{AC} \cdot \overrightarrow{HB}$。

在射影几何中，问题的背景是调和比理论，亦即它可归结为证 $(AB, HC) = -1$。求解方法要比初等方法简洁。解法如下：

若过 B 作 AM 的平行线交 MH、MC 或其延长线于 D、E，则 $\angle DMB = \angle CMB$，又因为 $\angle AMB$ 为直角，即 $AM \perp BM$。所以，$MB \perp DE$，从而 $DB = BE$。

考虑直线 DE 上的无穷远点 p_∞，则

$(DE, BP_\infty) = -1$。

视 MD、ME、MB、MA 为一线束，此线束被直线 DE 所截得 D、E、B、P_∞，由中心投影下交比的不变性知

$$(MD\quad ME, MB\quad MA) = (DE, BP_\infty) = -1,$$

又线束 MD、ME、MA、MB 被直线 AC 所截，则

$$(HC, AB) = (MD\quad ME, MA\quad MB) = -1,$$

故$\overline{AH} \cdot \overline{BC} = \overline{AC} \cdot \overline{HB}$。

上述解答中，MA、MB、MH、MC 的调和性还可用以下方法证明：

注意到 MB 是 $\angle HMC$ 的平分线，而 MA 是 $\angle HMC$ 的外角平分线。若在 $\angle HMC$ 内取一点为坐标原点，MH、MC 的法式方程分别为 $\alpha = 0$，$\beta = 0$，则 MB、MA 的方程依次为 $\alpha - \beta = 0$ $\alpha + \beta = 0$。从而

$$(MD\quad ME, MA\quad MB) = -1。$$

射影几何中，交比是一种活力，它可将点列与线束不断转化，从而简化解题过程，本题就是一例。

例 4　射影到无穷远。

设三直线 P_1P_2、Q_1Q_2、R_1R_2 交于一点 S，P_1P_2、Q_1Q_2、R_1R_2 分别交两直线 OX_1、OX_2 于 P_1、Q_1、R_1 与 P_2、Q_2、R_2。求证 P_1Q_2 与 P_2Q_1 的交点，Q_1R_2 与 Q_2R_1 的交点，R_1P_2 与 R_2P_1 的交点在一直线上，且所在的直线通过点 O（图 3.13(1)）。

这是点线结合关系的问题，属射影几何的讨论对象，故可将直线 OS 投射到无穷远。这只需在直线 OX_1、OX_2 所定的平面 π 外任取一点 V，取平面 π' 平行于 V 与 OS 所定的平面，则以 V 为射影中心建立的 π 向 π' 的中心投影将 OS 射影成 π' 的无穷远直线 $O'_\infty S'_\infty$。

设 P_1'、Q_1'、R_1'、P_2'、Q_2'、R_2' 分别为 P_1、Q_1、R_1，P_2、

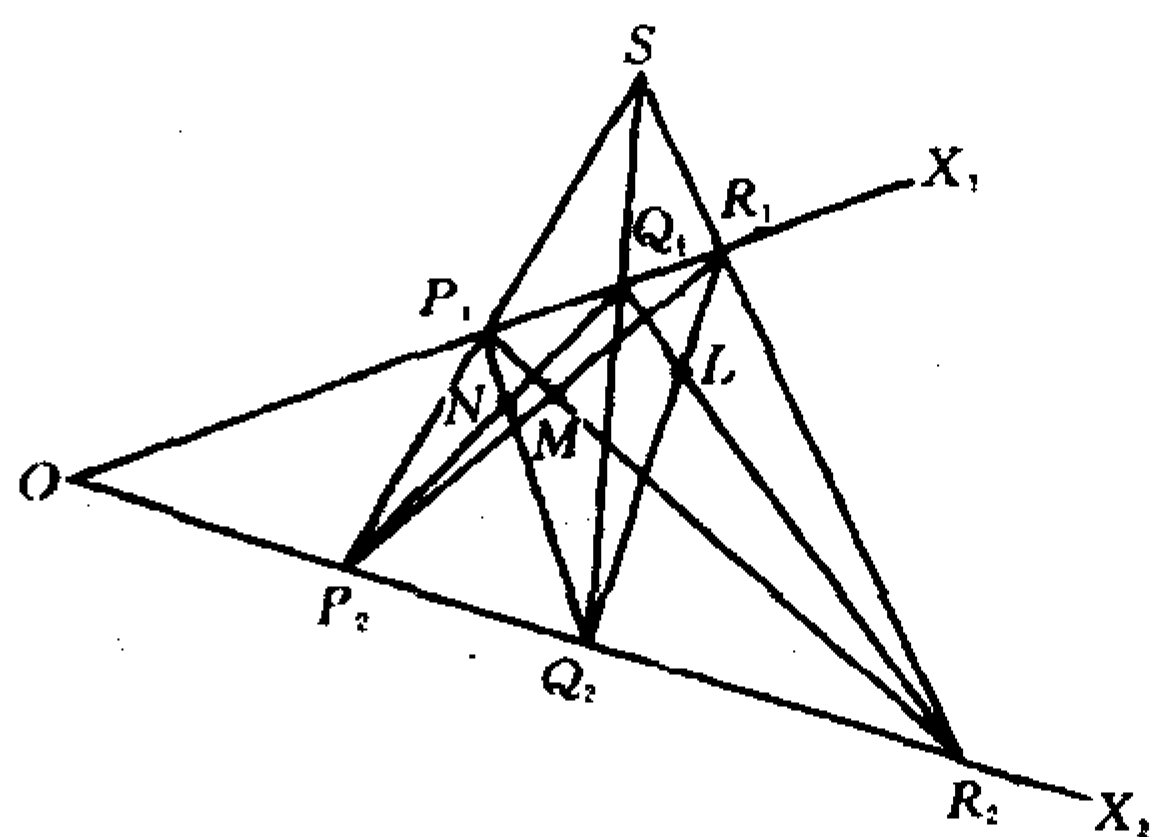


图 3.13(1)

Q_2、R_2 在此中心射影下的象。令

$$Q_1R_2 \times Q_2R_1 = L, P_1R_2 \times P_2R_1 = M,$$

$$P_1Q_2 \times P_2Q_1 = N。$$

则这三点在中心射影下的象为

$$Q_1'R_2' \times Q_2'R_1' = L', P_1'R_2' \times P_2'R_1' = M',$$

$$P_1'Q_2' \times P_2'Q_1' = N'。$$

由于

$$P_1'P_2' \parallel Q_1'Q_2' \parallel R_1'R_2', P_1'Q_1'R_1' \parallel P_2'Q_2'R_2',$$

故 $L'M'N' \parallel P_1'Q_1'R_1' \parallel P_2'Q_2'R_2'$,

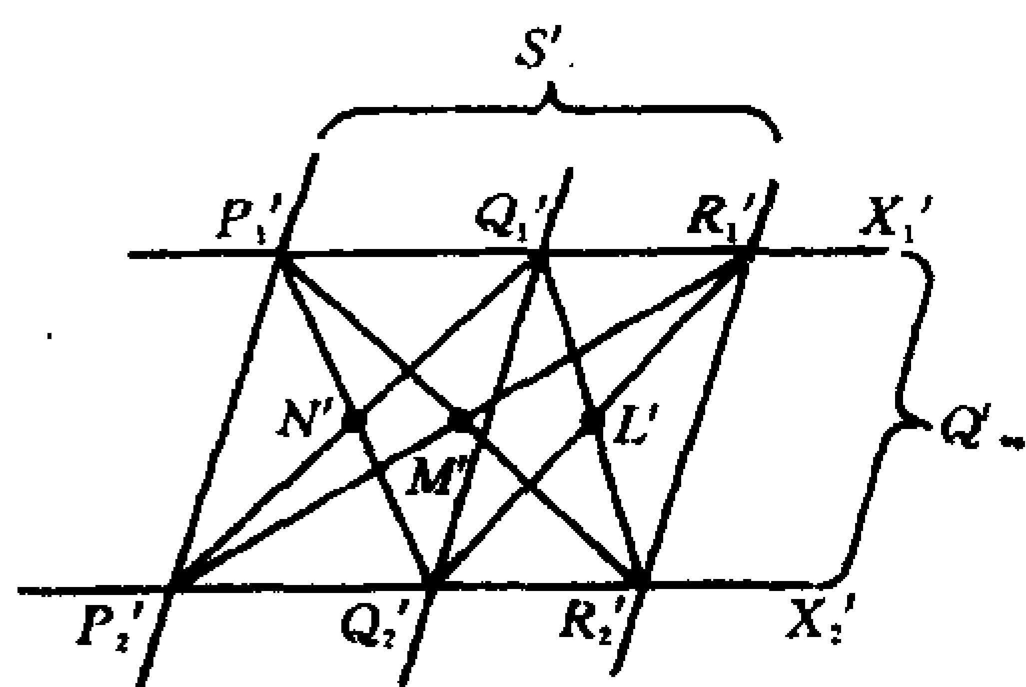


图 3.13(2)

亦即这三直线交于 O'_∞。由于中心射影保持点线结合关系不变,故 L、M、N 共线,且通过定点 O(图 3.13(2))。

例 5　高观点之下的中学几何竞赛题。

如图 3.14(1)，设 CD 上有一点 B，直线 $AC \parallel EB$，$AB \parallel ED$，则 $\triangle AEB$ 面积是 $\triangle ABC$ 面积和 $\triangle EDB$ 面积的比例中项。

这是武汉市 1978 年中学数学竞赛试题。从射影几何的观点考察，它是以平行投影为背景，从如下的问题中引发出来的。其问题是：

设 $C'D'$ 上有一点 B'，以 $D'B'$、$B'C'$ 为边在同侧分别作正三角形 $D'B'E'$ 和 $A'B'C'$（图 3.14(2)）。若 $B'C' = 1$，$D'B' = a$，则此时 $A'C' \parallel E'B'$，$A'B' \parallel E'D'$。且

$$S_{\triangle A'B'C'} = \frac{\sqrt{3}}{4},S_{\triangle D'E'B'} = \frac{\sqrt{3}}{4}a^2,S_{\triangle A'B'E'} = \frac{\sqrt{3}}{4}a。$$

从而必有

$$S^2_{\triangle A'B'E'} = S_{\triangle A'B'C'} \cdot S_{\triangle B'E'D'}。$$

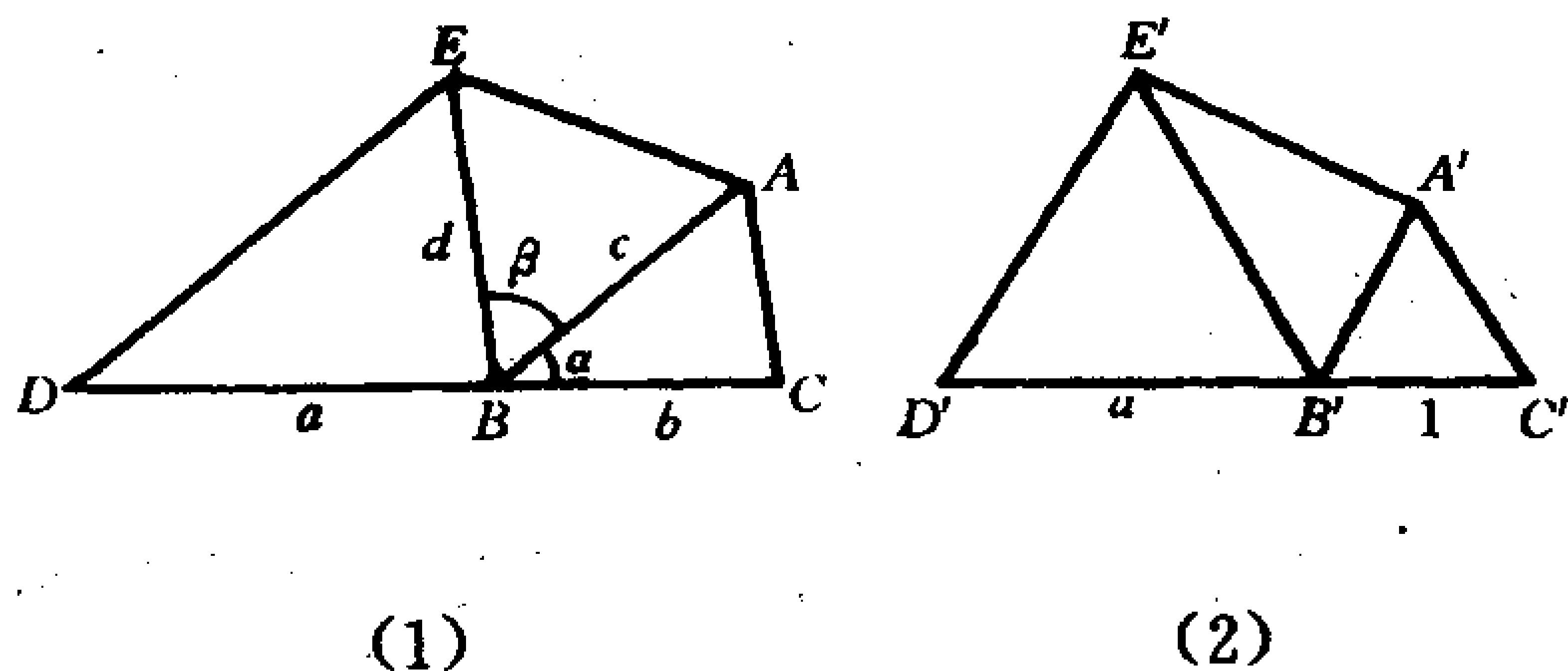


图 3.14

由于平行投影不改变两线段的平行关系，即平行线段经平行投影后仍为平行线段，且这种投影又不改变两图形的面积比。同时，任一三角形经一定的平行投影后总可变为正三角形。既然

$$S^2_{\triangle A'B'E'} = S_{\triangle A'B'C'} \cdot S_{\triangle E'B'D'}$$

对(2) 图成立，则对图(1) 必有

$$S^2_{\triangle ABE} = S_{\triangle ABC} \cdot S_{\triangle EBD}$$

成立，从而就有上述的竞赛试题。

竞赛试题的设计过程，为我们求解竞赛题提供了宝贵信息，即图(1)的问题可用计算方法证明，其证法如下：

设 $BC = b, BA = c, BE = d, \angle ABC = \alpha, \angle EBA = \beta$，则

$$S_{\triangle ABC} = \frac{1}{2}bc\sin\alpha, S_{\triangle ABE} = \frac{1}{2}cd\sin\beta。$$

又 $\quad \dfrac{d}{\sin\alpha} = \dfrac{DB}{\sin\beta}$，故 $DB = \dfrac{d\sin\beta}{\sin\alpha}$，从而

$$S_{\triangle BED} = \frac{1}{2}d^2\frac{\sin\beta}{\sin\alpha}\cdot\sin(\alpha+\beta)。$$

注意到 $b\sin(\alpha+\beta) = c\sin\beta$，

$$\therefore \quad S^2_{\triangle ABE} = S_{\triangle ABC}\cdot S_{\triangle DBE}。$$

3.7 排序、迭代和有向化

1. 序关系和排序

(1) 序关系与大小关系。

序关系是人们生活中的常识，又是数学研究的对象。在现代数学中，序结构与代数结构、拓扑结构一起构成了数学的三大基础结构，结构主义学派，即布尔巴基学派称之谓母结构。

一个集合 E 叫做有序或全序的是指，如果在 E 的元素之间定义了一个关系“$\prec$”，且具有以下性质：

1° 自反性：若 $x \in E$，则 $x \prec x$；

2° 反对称性：若 x、$y \in E$，$x \prec y$，且 $y \prec x$，则 $x = y$；

3° 传递性：若 x、y、$z \in E$，$x \prec y$，又 $y \prec z$，则 $x \prec z$；

4° 可比性：若 x、$y \in E$，$x \neq y$，则必有或者 $x \prec y$，或

者 $y \prec x$。

例如，既约的正的真分数集 M 按大小关系"$\leqslant$"符合上述四条公理。又同样可以验证既约真分数集，按如下的字典排列也可构成有序集：

设 $r_1 = \frac{p_1}{q_1}, r_2 = \frac{p_2}{q_2}$。若 $q_1 < q_2$，或者 $q_1 = q_2$ 且 $p_1 \leqslant p_2$，则称 r_1 小于或等于 r_2，记为 $r_1 \prec r_2$。

按这种顺序排列，既约的正的真分数，依次为：

$$\frac{1}{2},\frac{1}{3},\frac{2}{3},\frac{1}{4},\frac{3}{4},\frac{1}{5},\frac{2}{5},\frac{3}{5},\frac{4}{5},\frac{1}{6},\frac{5}{6},\frac{1}{7},\cdots$$

这说明同一集合中的元素可按不同方法排定次序。

又如，平面上的点亦可排成全序。设任给两点 (x_1, y_1)，(x_2, y_2)。约定若 $x_1 < x_2$，或者 $x_1 = x_2$ 且 $y_1 \leqslant y_2$，记为 $(x_1, y_1) \prec (x_2, y_2)$。亦即平面上点的次序可以先比第一坐标，第一坐标相等时再比第二坐标。显然，"$\prec$"满足四条全序公理：

1°$(x, y) \prec (x, y)$；

2° 若 $(x_1, y_1) \prec (x_2, y_2)$，$(x_2, y_2) \prec (x_1, y_1)$，则由定义必有 $x_1 = x_2, y_1 = y_2$，即 $(x_1, y_1) = (x_2, y_2)$。

3° 设 $(x_1, y_1) \prec (x_2, y_2)$，$(x_2, y_2) \prec (x_3, y_3)$，则由第一式知 $x_1 < x_2$ 或 $x_1 = x_2$ 且 $y_1 \leqslant y_2$。又由第二式知 $x_2 < x_3$ 或者 $x_2 = x_3$ 且 $y_2 \leqslant y_3$。组合诸情况有 $x_1 < x_3$，或者 $x_1 = x_3$，且 $y_1 \leqslant y_3$。亦即 $(x_1, y_1) \prec (x_3, y_3)$。

4° 任给 (x_1, y_1)、(x_2, y_2) 且 $(x_1, y_1) \neq (x_2, y_2)$，则必有 $x_1 \neq x_2$，或者 $x_1 = x_2$ 且 $y_1 \neq y_2$。若 $x_1 < x_2$，则必 $(x_1, y_1) \prec (x_2, y_2)$；若 $x_1 > x_2$，便有 $(x_2, y_2) \prec (x_1, y_1)$；若 $x_1 = x_2$ 且 $y_1 \leqslant y_2$，则 $(x_1, y_1) \prec (x_2, y_2)$；若 $x_1 = x_2$　且 $y_1 > y_2$，则 $(x_2, y_2) \prec (x_1, y_1)$，即 (x_1, y_1) 和 (x_2, y_2) 是可比

的。

因此，平面上点已排成了全序。由于复数与平面上点构成一一对应，故复数也排成全序，其序关系是：实部小的复数在前，实部相等，虚部小的在前。

序关系不等于大小关系。一个序关系只在与四则运算有协调关系时才称为大小，这种协调关系的特征具有如下的性质：

5° 加法保序性：若 $a \geqslant b$，对任何 c 有 $a+c \geqslant b+c$；

6° 乘正数保序性：若 $a \geqslant b$，对任何 $c \geqslant 0$，有 $ac \geqslant bc$。亦即一个序关系当满足 5°、6° 时称为数目次序，这时元素间有大小关系。

既约的正的真分数按字典排列的顺序不是大小关系，因为它不满足上述两条保序性。例如，

$$\frac{1}{2} \prec \frac{1}{3},$$

但 $$\frac{1}{2}+\frac{1}{6}=\frac{2}{3} \succ \frac{1}{2}=\frac{1}{3}+\frac{1}{6}。$$

又 $$\frac{1}{8} \prec \frac{1}{9},$$

但 $$\frac{1}{8} \times 3=\frac{3}{8} \succ \frac{1}{3}=\frac{1}{9} \times 3。$$

复数无大小关系。即无论怎样定义元素间的关系，都不能同时适合 1°—6°。对此，可用反证法证明。

假定复数存在着关系 $\prec$ 满足 1°—6°，则必将引起矛盾。这只要考察元素 0 与 i：

因为 $0 \neq i$，所以或者 $0 \prec i$，或者 $i \prec 0$。

① 若 $0 \prec i$，由 6° 知 $0 \cdot i \prec i \cdot i$ 成立，亦即 $0 \prec -1$。再由 6° 得 $0 \cdot i \prec (-1) \cdot i$，即 $0 \prec -i$。又由 5°，$0+i \prec (-i)+i$ 成立，即 $i \prec 0$ 矛盾。

② 若 $i \prec 0$，由 5° 知 $i + (-i) \prec 0 + (-i)$ 成立，亦即 $0 \prec -i$。又由 6° 得 $0 \cdot (-i) \prec (-i)(-i)$ 亦即 $0 \prec -1$，再由 6° 有 $0 \cdot (-i) \prec (-1)(-i)$ 得 $0 \prec i$ 矛盾。

由 ①、② 知，复数无大小关系。

对序关系，有时定义满足序公理中前三条，亦即具有自反性、反对称性、传递性的集合称为半序集。例如，自然数以整除为半序。若设 N 为自然数集，a、$b \in N$，a 整除 b（记以 $a|b$）定义为 $a \prec b$。不难验证这是半序。

因为 $a|a$；若 $a|b$，$b|a$，则 $a = b$；若 $a|b$，$b|c$，则 $a|c$。故它满足半序公理的三条。但任何两整数之间未必有整除关系，因而没有可比性，故这样的关系只能构成半序，不能为全序。

函数依全局大小为半序。

设 F 是闭区间 $[a,b]$ 上一切实函数全体，$f(x)$、$g(x) \in F$，若对 $x \in [a,b]$ 有 $f(x) \leqslant g(x)$，则规定 $f \prec g$，亦即 $g(x)$ 在 $[a,b]$ 中各个 x 处全面地大于、至少等于 $f(x)$，这时 F 按此也构成半序集。由于任何两个函数之间不一定都存在"全面优势"，所以 F 不为全序集。

如果说实数集的主要特征之一是全序，那么实函数全体所成之集的特征之一就是半序。从全序到半序，反映了人们对客观世界数量关系的认识又深入了一步。

(2) 排序与排序原理。

排序是这样一种思想方法，即在研讨或处理问题时，根据需要把问题中所涉及的量用某种方式排定次序，然后借助构造的序关系再行分析、解决问题。

例如，设有 7 个点，其中任何 3 点不在一直线上，任何 4 点不在一圆周上，问是否一定存在过 3 点的圆周，使其余的点中的一半点在圆内，一半点在圆外？

据题意，总存在某两点的一条直线，使其余 5 点均在这直线的同侧。若这两点为 A、B，其余 5 点为 $P_i(i=1,\cdots,5)$，将这 5 点按 $\angle AP_iB$ 的大小顺序排定为：

$$\angle AP_1B<\angle AP_2B<\angle AP_3B<\angle AP_4B<\angle AP_5B。$$

这时，过 A、B、P_3 作圆，则 P_1、P_2 必在圆外，P_4、P_5 必在圆内。这里将 P_i 按 $\angle AP_iB$ 的大小顺序排定次序是求解的关键。因为只有当这种次序排定后，我们才能具体地寻找适合问题的解。亦即排序往往可使某些一般情形的问题化归为某些特殊情况的问题。这种按需要给对象排定次序的思想就是排序。

应用排序思想，类似地还可证明：平面上 $2n+3$ 个点中任何 3 点不共线，任何 4 点不共圆，则一定存在过其中 3 点的圆周，使得其余 $2n$ 个点的一半在圆内，另一半在圆外。这时，当选定两点 A、B 使其余 $2n+1$ 个点均在 AB 连线同侧后，要对 $P_i(i=1,2,\cdots,2n-1)$ 按 $\angle AP_iB$ 的大小排序。

应用排序思想，还可得到数学上的一个重要原理。这是指：

设 $a_1,a_2,\cdots,a_n$ 和 $b_1,b_2,\cdots,b_n$ 是 $2n$ 个非负实数，而 $b_1^*,b_2^*,\cdots,b_n^*$ 是 $b_1,b_2,\cdots,b_n$ 的任一排列，

$$S=a_1b_1^*+a_2b_2^*+\cdots+a_nb_n^*。$$

则在和数 S 中，最大的是 a_i 和 b_i 按同顺序相乘、相加所得的数，最小的是 a_i 和 b_i 按反顺序相乘相加所得的数。事实上，设

$$S'=a_1b_1^*+\cdots+a_ib_j^*+\cdots+a_jb_i^*+\cdots+a_nb_n^*,$$

又 $a_i<a_j$，则

$$S-S'=(a_ib_i^*+a_jb_j^*)-(a_ib_j^*+a_jb_i^*)$$

$$= (a_i - a_j)(b_i^* - b_j^*)\begin{cases} > 0, b_i^* < b_j^* \text{ 时}, \\ < 0, b_i^* > b_j^* \text{ 时}。\end{cases}$$

由于给定的数a和b只能构成有限个不同的和数，在它们中总有最大的和最小的，而当a_i或$b_i(i=1,2,\cdots,n)$都相等时，取等号，故而结论成立。

排序原理有着重要应用，许多著名不等式都可由它推出。如，对n个正数$x_1,\cdots,x_n$，有

$$\sqrt[n]{x_1 \cdot x_2 \cdot \cdots \cdot x_n} \leqslant \frac{x_1 + x_2 + \cdots + x_n}{n}。$$

这只要在$2n$个数中令$c = \sqrt[n]{x_1x_2\cdots x_n}$，且

$$a_1 = \frac{x_1}{c}, a_2 = \frac{x_1x_2}{c^2}, \cdots, a_n = \frac{x_1 \cdot \cdots \cdot x_n}{c^n} = 1,$$

$$b_1 = \frac{1}{a_1}, b_2 = \frac{1}{a_2}, \cdots, b_n = \frac{1}{a_n}。$$

注意到a_i、$b_i(i=1,2,\cdots,n)$反序，便有

$$\begin{aligned} n &= \underbrace{1 + \cdots + 1}_{n\text{个}} \\ &= a_1b_1 + \cdots + a_nb_n \\ &\leqslant a_1b_n + a_2b_1 + a_3b_2 + \cdots + a_nb_{n-1} \\ &= \frac{x_1}{c} + \frac{x_2}{c} + \cdots + \frac{x_n}{c}。\end{aligned}$$

等号仅在$\frac{x_1}{c} = \frac{x_1}{c} \cdot \frac{x_2}{c} = \cdots = \frac{x_1}{c} \cdot \frac{x_2}{c} \cdots \frac{x_n}{c} = 1$，或$x_1 = x_2 = \cdots = x_n = c$时成立。

此不等式称为算术-几何不等式，法国数学家柯西用数学归纳法解答了此题，故又称它为柯西不等式。

许多实际生活中的优化问题，也可用排序原理来解决。例如，1978年全国数学竞赛有如下一个问题，题目是：

有10人各自拿一水桶同时到某一自来水龙头上盛

水。已知各人水桶的容量不等，问怎样安排打水的顺序才能使花费的总时间为最少？

由于第一、第二、…、第十人打水时等待的人数分别为10人，9人，…，1人。同时，因各人水桶的容量不等，故它们分别盛满水桶的时间也不同，设分别为$t_1,t_2,\cdots,t_{10}$，易知盛水花费的总时间为：

$$10t_1+9t_2+\cdots+2t_9+t_{10},$$

由于10，9，…，1是倒序，故按排序原理只有当$t_1,t_2,\cdots,t_{10}$是由小到大即顺序时才能使$10t_1+9t_2+\cdots+2t_9+t_{10}$是反序和，而反序和是最小的。因此，盛水花费的总时间最少的方法是按桶的容积大小从小到大的次序逐个盛水。

排序不仅对数学，而且对任何一种理性的认识活动都具有重要的意义。因为任何思维都包括着信息的获取、储存、加工和输出等过程，其中对信息的加工处理过程，除了要对信息进行核实外，分类、排序是不可缺少的。例如，通过对信息的分析、比较才能排出最重要的、重要的、次重要的、不重要的、可忽略的，外表的、表里兼有的、及里的，等等。没有对信息的分类和排序，任何科学的抽象都是不可能的。

2. 迭代与计算机

迭代是一种数值方法。例如，求方程$x=g(x)$的解，可选择适当的初始值x_0，得

$$x_1=\psi(x_0),x_2=\psi(x_1),\cdots,x_{n+1}=\psi(x_n),\cdots$$

在一定条件下，若x_n有极限，则x_{n+1}也有相同极限，亦即是原方程的根。这种通过逐次代入同一关系式而实现问题解的方法即是迭代。

迭代这一计算方法自计算机出现后应用更加广泛，

因为迭代的特点是逐次代入同一关系式，亦即机械地重复同一动作，这正是计算机的长处。

例如，要计算多项式 $f(x)$ 在 $x=8$ 时的值，其中

$$f(x)=Ax^6+Bx^5+Cx^4+Dx^3+Ex^2+Fx+G。$$

若改写 $f(x)$ 为

$$f(x)=((((((Ax+B)x+C)x+D)x+E)x+F)x+G,$$

于是，数值计算化为这样两个动作的重复进行：自第一系数 A 起，依 A、B、C、D、E、F、G 的顺序，乘 x 加下一系数，直至取完最后一个系数为止。显然，这样计算 $f(x)$ 的值对于计算机来说，是很容易的。

又如，一个正数 a，求它的平方根，即是求方程

$$x^2-a=0 \quad 或 \quad x=\frac{1}{2}\left(x+\frac{a}{x}\right)$$

的正根。用迭代法，考虑迭代公式

$$x_{n+1}=\frac{1}{2}\left(x_n+\frac{a}{x_n}\right)。$$

取 $x_0>0, n=0,1,2,\cdots$

注意到 $x_n>0$，显然 x_n 是有下界的。又

$$\frac{x_{n+1}}{x_n}=\frac{1}{2}\left(1+\frac{a}{x_n^{\ 2}}\right),$$

且

$$x_n=\frac{1}{2}\left(x_{n-1}+\frac{a}{x_{n-1}}\right)$$

$$=\sqrt{a}\cdot\frac{1}{2}\left(\frac{x_{n-1}}{\sqrt{a}}+\frac{\sqrt{a}}{x_{n-1}}\right)\geqslant\frac{\sqrt{a}}{2}\cdot 2=\sqrt{a},$$

故

$$\frac{x_{n+1}}{x_n}=\frac{1}{2}\left(1+\frac{a}{x_n^{\ 2}}\right)\leqslant\frac{1}{2}\left(1+\frac{a}{(\sqrt{a})^2}\right)=1。$$

$$\therefore \quad x_{n+1} \leqslant x_n。$$

即 x_n 为单调下降的，从而 x_n 收敛 c，$c = \frac{1}{2}\left(c + \frac{a}{c}\right)$，故

$$c = \sqrt{a}。$$

这一迭代的误差可作如下估计：

记 $\delta_n = x_n - \sqrt{a}$，则

$$x_n = \delta_n + \sqrt{a},$$

$$x_{n+1} = \frac{1}{2}\left(x_n + \frac{a}{x_n}\right) = \frac{1}{2}\left[(\delta_n + \sqrt{a}) + \frac{a}{\delta_n + \sqrt{a}}\right]$$

$$= \frac{\delta_n^2 + 2\sqrt{a}\,\delta_n + 2a}{2(\delta_n + \sqrt{a})},$$

$$\delta_{n+1} = x_{n+1} - \sqrt{a} = \frac{\delta_n^2 + 2\sqrt{a}\,\delta_n + 2a}{2(\delta_n + \sqrt{a})} - \sqrt{a}$$

$$= \frac{\delta_n^2}{2(\delta_n + \sqrt{a})} \leqslant \frac{\delta_n^2}{2\sqrt{a}}。$$

本章第 5 节中谈到的黄金数

$$\omega = \frac{-1 + \sqrt{5}}{2} \approx 0.6180339,$$

它在优选法中有极为重要的应用。ω 是二次方程 $\omega^2 + \omega - 1 = 0$ 的正根。现用迭代法求其近似根。改写原二次方程为 $x = \frac{1}{x+1}$，取迭代公式

$$x_{n+1} = \frac{1}{x_n + 1}。$$

令 $f(x) = x^2 + x - 1$，且 $f(0) = -1 < 0$， $f(1) = 1 > 0$，

故　$f(x)$ 在 $(0,1)$ 间有解 x^*。

取 $x_0 = 0$ 代入迭代公式得 $x_1 = 1$（1 次迭代），继续这样的过程，得

$$x_2 = \frac{1}{1+x_1} = \frac{1}{2}$$(2 次迭代)，$x_3 = \frac{1}{1+x_2} = \frac{2}{3}$(3 次迭代)，$x_4 = \frac{3}{5}, x_5 = \frac{5}{8}, x_6 = \frac{8}{13}, \cdots$。

得到了解的一串近似值

$$\frac{1}{1}, \frac{1}{2}, \frac{2}{3}, \frac{3}{5}, \frac{5}{8}, \frac{8}{13}, \cdots$$

可记为：

$$x_1 = \frac{F_1}{F_2}, x_2 = \frac{F_2}{F_3}, x_3 = \frac{F_3}{F_4}, \cdots, x_n = \frac{F_n}{F_{n+1}},$$

其中

$$F_1 = F_2 = 1, F_n = F_{n-2} + F_{n-1}, n > 2。$$

可证明它收敛于$\frac{\sqrt{5}-1}{2} \approx 0.618$。其中，数列$\{F_n\}$即是著名的斐波纳契数列。在优选法中，上述近似分数常用来安排试验点，即是优选法中的分数法。

如上迭代时获得的著名数列$\{F_n\}$，在自然界具有奇妙的性质。比如，植物的叶子在茎上的排列，同一种植物具有一定规律。若称位于茎周同一母线位置的两片叶子为一周期，又记

W = 每个周期叶子围绕的圈数 / 每个周期中的全部叶子数。

则 W 将是一些特定的数，仅随植物品种不同而异。如

$$W_{榆树} = \frac{1}{2}, W_{山毛榉} = \frac{1}{3}, W_{樱桃} = \frac{2}{5}, W_{梨树} = \frac{3}{8},$$

$$W_{柳树} = \frac{5}{13}, \cdots$$

这串分数的分子和分母都各自组成了一个斐波纳契数列。类似地，$\{F_n\}$ 还与兔子的繁殖、雄蜂的家族、钢琴琴键排列等问题密切相关。

3. 有向化方法

所谓有向化是指某些问题中的量或某些问题本身，原无方向的意义，但处理问题时，按需要把它们看成有方向的，这样的思想称为有向化。例如，对距离、面积、体积等规定方向，就产生了有向距离、有向面积和有向体积。又如，某系统由 n 个量决定，有时我们就把这个系统看作一些有方向的量的集合，这些有方向的量均可沿 n 个方向自由变动等。

有向化的突出作用是：

(1) 它能将彼此孤立的某些量，用方向有机结合起来，使之成为一个相互关联的、和谐的统一整体。

例如，若干线段相加减，如果我们把所论线段都赋予方向的意义，那么原问题就统一为求首尾相接的同一直线上的若干有向线段的和。同样，平面上若干角的加、减，把角看成是有方向的，也可化为求一列系有向角的和：

$$\measuredangle(a,b)+\measuredangle(b,c)+\cdots+\measuredangle(l,m)+\measuredangle(m,n)$$
$$=\measuredangle(a,n)。$$

其中，有向角 $\measuredangle(a,b)$ 是指，任意一点 O，过 O 分别作 a、b 两直线的平行线 l、l'，则将 l 绕 O 点旋转重合于 l' 的角称为 a、b 的有向角，记为 $\measuredangle(a,b)$。通常旋转方向逆时针的角值为正，顺时针的角值为负。

平面极坐标，若用有向量描述，便为：

对一固定点 O——极点，自 O 引一固定轴 OX——极轴，平面上的任一点 P，OP 所在直线为 l，$\vec{e}$ 为 l 上的单位向量，$\psi=\measuredangle(\overrightarrow{OX},\vec{e})$，$\overrightarrow{OP}=\rho\vec{e}$，则$(\rho,\psi)$ 叫做点 P 的极坐标。

又如，几何中的某些不同概念和定理，应用方向量处

理，就可得到统一的描述。如用有向线段可统一叙述内、外位似，梅奈劳斯定理和齐瓦定理等。用有向角，圆中 A、B、C、D 四点，不论次序如何，其共圆的充要条件都为：

$$\measuredangle ACB = \measuredangle ADB \neq 0$$

（$\measuredangle ACB$ 是指直线 CA 和 CB 间的有向角）（图 3.15）。

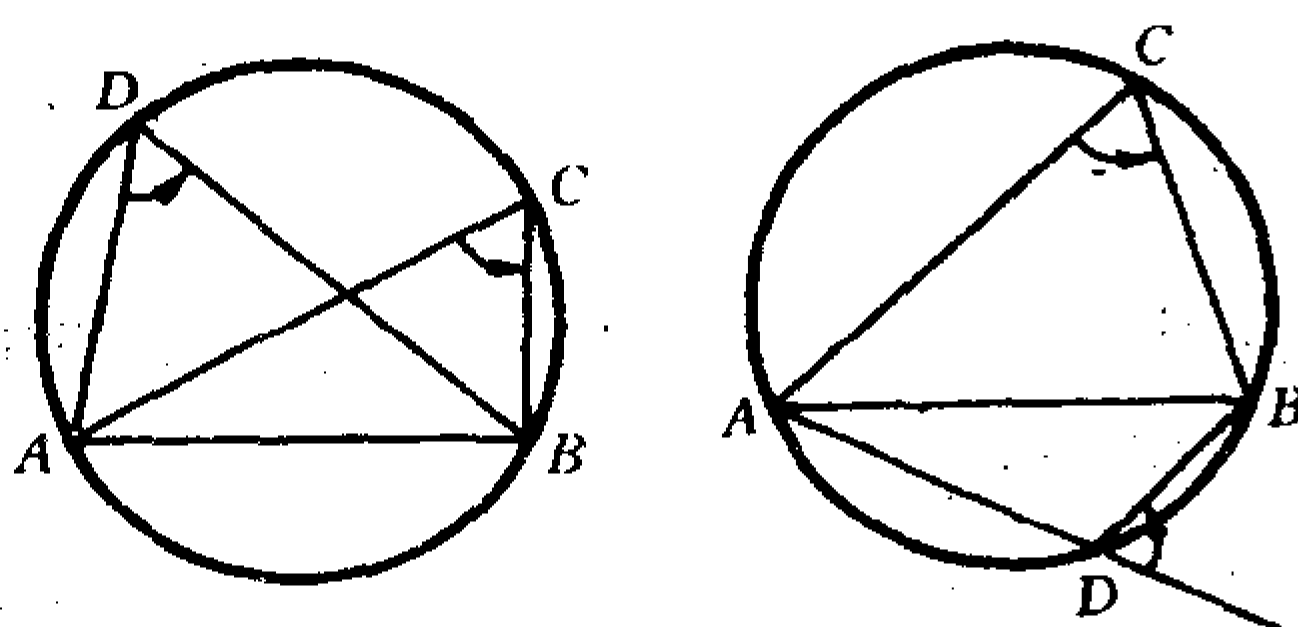


图 3.15

同时，用有向化处理某些问题，常常可以取得内容和形式的和谐一致。

例如，平面上点对于直线或空间中点对于平面的位置关系，应用有向化处理，可得到有向距离（离差）的概念。而这有向距离，可用点的坐标代入直线或平面的法式方程的左端得到。应用有向距离，要断定给定的点是否是凸多面体内部的点，就只要对所论多面体的任一个面，考虑它不在这个面的其余各顶点与给定点相对于这平面的离差是否都有相同的符号，若都相同（不为零），则该已知点必是多面体内部的点。这里，点的坐标代入直线和平面的法式方程这一形式与有向距离的内容之间取得了和谐的一致。

又如，直角坐标系中，有向面积、体积为

$$\overline{S}_{\triangle A_1A_2A_3} = \frac{1}{2}\begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix},$$

$$\overline{V}_{\text{四面体}A_1A_2A_3A_4}=\frac{1}{6}\begin{vmatrix} x_1 & y_1 & z_1 & 1 \\ x_2 & y_2 & z_2 & 1 \\ x_3 & y_3 & z_3 & 1 \\ x_4 & y_4 & z_4 & 1 \end{vmatrix}。$$

其中点 A_i 的平面坐标为(x_i,y_i),空间坐标为(x_i,y_i,z_i)。面积、体积的量值,不仅与顶点坐标的值有关,而且与它们在行列式中排列的顺序有关。行列式的表达形式与有向面积、体积的内容取得了完美的统一。

(2) 有向化方法为解题提供了统一处理的方法,省去了许多分情况讨论。

数学中的繁琐、复杂,莫过于证明或解题中存在情况众多的讨论。英国数学家哈代(Hardy) 曾指出:"我们不希望数学定理的证明中分很多'情况',实际上'列举情况'是数学论证中一种比较呆板的'形式',一个数学证明应该像一个简单而轮廓分明的星座,而不是银河系中杂乱无章的星群。"[33] 由于有向化方法可将各种片面、孤立的情况联结成为一个统一的整体,这样就为研讨问题提供了统一处理的方法,从而避免了求解和论证问题时出现繁琐、复杂的分情况讨论,因而用这样的方法处理问题,必然使解题方法具有高度的简洁性和概括性。

例如,在 $\triangle ABC$ 三边 BC、CA、AB 所在直线上各取异于顶点的点 X、Y、Z,则 $\odot AYZ$、$\odot BZX$、$\odot CXY$ 共点。

若不用有向角,就要区分 X、Y、Z 三点在边上或边的延长线上的诸情况(图 3.16 仅画出其两种)。用有向角处理,便可概括诸多情况的证明于一,方法是:

设 $\odot BZX$ 与 $\odot CXY$ 的第二交点为 O,连 OX、OY、OZ,则

$$\measuredangle AYO=\measuredangle CXO=\measuredangle BZO=\measuredangle AZO,$$

故　$\measuredangle AYO = \measuredangle AZO$，从而$O$在$\odot AYZ$上，即三圆共点。

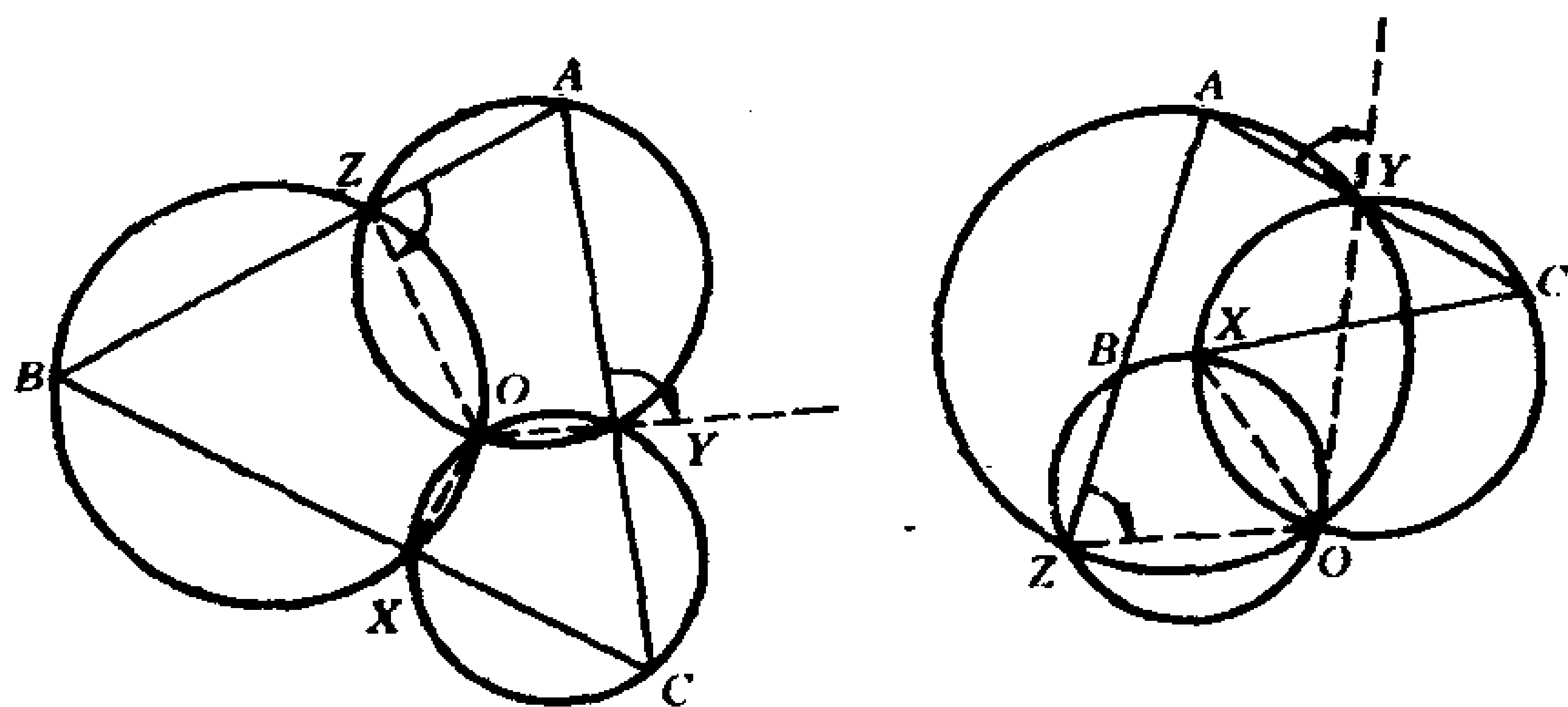
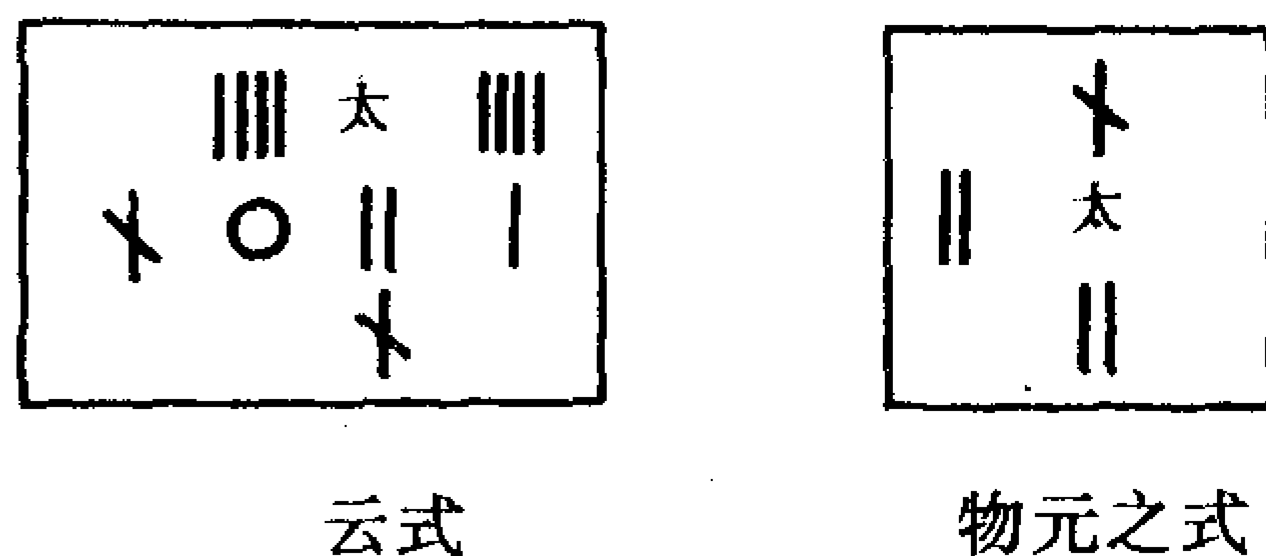


图 3.16

方向、位置在中国古算中是很突出的。例如，金元时期，中算家发明了天元符号，创立了天元术，在天元术基础上又发展出四元术。对此，《四元玉鉴》元莫若序中记述道："以元气居中，立天元一于下，地元一于左，人元一于右，物元一于上，阴阳升降，进退左右，互通变化，错综无穷。"即是说，"太"居中表常数项，天、地、人、物四个未知数分居"太"的下、左、右、上四个位置，诸元的幂次由它们与"太"的位置决定，相邻两元的幂次之积记入相应行、列的交叉处，不相邻诸元之积证入夹缝中。如朱世杰《四元玉鉴·假令》"四象会元"中的"云式"和"物元之式"分别表示：

$-x^2+2x-xy^2+4y+xz+4z$ 和 $2x+2y-u$（图 3.17）。

这种用方向、位置来表达未知数的方法出现在金元时期，在当时确是一个很大的进步，但这种符号后来没有继续进步和向前发展，至 16 世纪后欧洲出现了符号代数，相比之下，我国的方向、位置性质符号就显得落后了。



云式　　物元之式

图 3.17

3.8 回归分析与马尔可夫预测

1. 回归分析思想与一元线性回归

现实世界中事物的运动都是和它周围的事物互相联系、相互影响的，因而反映其运动的变量之间就必然存在着一定的关系。这些变量之间关系的发现，通常是通过试验或实验测定得到一批数据，然后再行分析处理，找出其相互关系。回归分析就是研究变量之间这种相关关系的一种数理统计方法。它可以根据观察数据，确定变量之间是否具有某种相关关系，并给出近似的数学表达式。据此，还可对某变量进行预测或控制。

一元线性回归是处理两个变量相关关系的最简单的回归分析，其基本思想是：设有 n 对观察数据

$x: x_1, x_2, \cdots, x_n$;

$y: y_1, y_2, \cdots, y_n$。

据此分别在直角坐标平面上作出 (x_i, y_i) 的点，这样的图称为散点图。若图中各点分布趋于一直线，这时就说变量之间存在着线性相关关系，并可用关系式 $y = \beta_0 + \beta_1 x$ 表示。其中 β_0、β_1 是需从已知数据进行估计的估计值，分别记以 a, b。则称 $\hat{y} = a + bx$ 为回归方程，a、b 为回归系数。又称 $\hat{y}_i = a + bx_i$ 为回归值，亦即 y_i 的估计值。$e_i = y_i$

$-\hat{y}_i$ 反映了估计值与观察值之间的偏差，称为残差。由于 n 个残差相加时可能因正负抵消，故其代数和不能真实地表示其总的偏离情况，又用残差的绝对值之和来表示又不利于运算，最简便的方法就是取残差的平方和，即

$$Q=\sum_{i=1}^{n} e_i^2=\sum_{i=1}^{n}(y_i-a-bx_i)^2,$$

并使其最小，这就是最小二乘法。待估计的 a、b 值应满足方程组：

$$\begin{cases}\dfrac{\partial Q}{\partial a}=-2\sum\limits_{i=1}^{n}(y_i-a-bx_i)=0\\[2ex]\dfrac{\partial Q}{\partial b}=-2\sum\limits_{i=1}^{n}(y_i-a-bx_i)x_i=0,\end{cases}$$

即

$$\begin{cases}\sum\limits_{i=1}^{n} y_i=na+b\sum\limits_{i=1}^{n} x_i\\[2ex]\sum\limits_{i=1}^{n} x_iy_i=a\sum\limits_{i=1}^{n}x_i+b\sum\limits_{i=1}^{n}x_i^2。\end{cases}$$

解得

$$a=\frac{1}{n}\sum_{i=1}^{n}y_i-\frac{b}{n}\sum_{i=1}^{n}x_i,$$

$$b=\left[\sum_{i=1}^{n}x_iy_i-\frac{1}{n}\left(\sum_{i=1}^{n}x_i\right)\left(\sum_{i=1}^{n}y_i\right)\right]\Bigg/\left[\sum_{i=1}^{n}x_i^2-\frac{1}{n}\left(\sum_{i=1}^{n}x_i\right)^2\right]。$$

记 $\bar{x}=\dfrac{1}{n}\sum\limits_{i=1}^{n}x_i,\ \bar{y}=\dfrac{1}{n}\sum\limits_{i=1}^{n}y_i,$

$$L_{xx}=\sum_{i=1}^{n}(x_i-\bar{x})^2=\sum_{i=1}^{n}x_i^2-\frac{1}{n}\left(\sum_{i=1}^{n}x_i\right)^2,$$

$$L_{xy}=\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})$$

$$=\sum_{i=1}^{n}x_iy_i-\frac{1}{n}\left(\sum_{i=1}^{n}x_i\right)\left(\sum_{i=1}^{n}y_i\right),$$

则 $a=\bar{y}-b\bar{x},\ b=L_{xy}/L_{xx}$。

由于 $\bar{x}$、$\bar{y}$、L_{xx}、L_{xy} 都可由观察值计算而得，故回归直线

便可求得。

注意到任何 n 对数据都可以用最小二乘法求出回归直线，这就需要考虑所求的回归直线是否可靠，因而就要对所求的回归直线进行显著性检验。通常用相关系数作为检 验统计量，以检验变量 x、y 之间是否具有直线关系。

我们把 $r = L_{xy}/\sqrt{L_{xx} \cdot L_{yy}}$ 称为变量 x 与 y 的相关系数。其中，

$$L_{yy} = \sum_{i=1}^{n}(y_i - \overline{y})^2 = \sum_{i=1}^{n} y_i^2 - \frac{1}{n}(\sum_{i=1}^{n} y_i)^2。$$

相关系数 r 满足不等式 $|r| \leqslant 1$。这是因为

$$\begin{aligned} Q &= \sum_{i=1}^{n}(y_i - a - bx_i)^2 = \sum_{i=1}^{n}(y_i - \overline{y} + b\overline{x} - bx_i)^2 \\ &= \sum_{i=1}^{n}(y_i - \overline{y})^2 - 2b\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) + b^2\sum_{i=1}^{n}(x_i - \overline{x})^2 \\ &= L_{yy} - \frac{2L_{xy}^2}{L_{xx}} + \frac{L_{xy}^2}{L_{xx}^2}L_{xx} = L_{yy}\left(1 - \frac{L_{xy}^2}{L_{xx}L_{yy}}\right) \\ &= L_{yy}(1 - r^2) \geqslant 0。\end{aligned}$$

由于 $L_{yy} \geqslant 0$，所以 $1 - r^2 \geqslant 0$，即 $|r| \leqslant 1$。当 $r > 0$ 时，称 x 与 y 正相关；当 $r < 0$ 时，称 x 与 y 负相关；当 $r = 0$ 时，称 x 与 y 不相关。当 $|r|$ 接近于 1 时，就认为 x 与 y 之间有明显的线性关系。

$|r|$ 怎样才算“接近”于 1，可以查相关系数检验表得临界值。若给定的显著性水平为 α(0.05 或 0.01)，则据 α 与对应的 $n - 2$($n - 2$ 是指 Q 中有 n 个变量，2 个约束条件，故其自由度为 $n - 2$)。查相关系数检验表，得临界值 $r_\alpha(n - 2)$。当 $|r| > r_{0.05}(n - 2)$ 时，就认为 y 与 x 的线性关系是显著的；当 $|r| > r_{0.01}(n - 2)$ 时，认为 y 与 x 的线性相关关系是高度显著的；而当 $|r| < r_{0.05}(n - 2)$ 时，则

认为 y 与 x 的线性相关关系不显著，在此情况下，求得的回归直线没有什么意义。

例 1　国民收入与财政收入的线性回归。

某市近几年国民收入与财政收入的统计资料如下：

国民收入(亿元)	4.6	3.5	5.0	6.4	8.3	8.9	9.0	9.5
财政收入(亿元)	0.4	0.5	0.7	1.1	1.6	1.8	1.9	2.2

据此要确定该市财政收入与国民收入之间的相关关系。

由表中 8 对值作出散点图，诸点分布趋于一直线(图 3.18)。据表计算有关数据，得

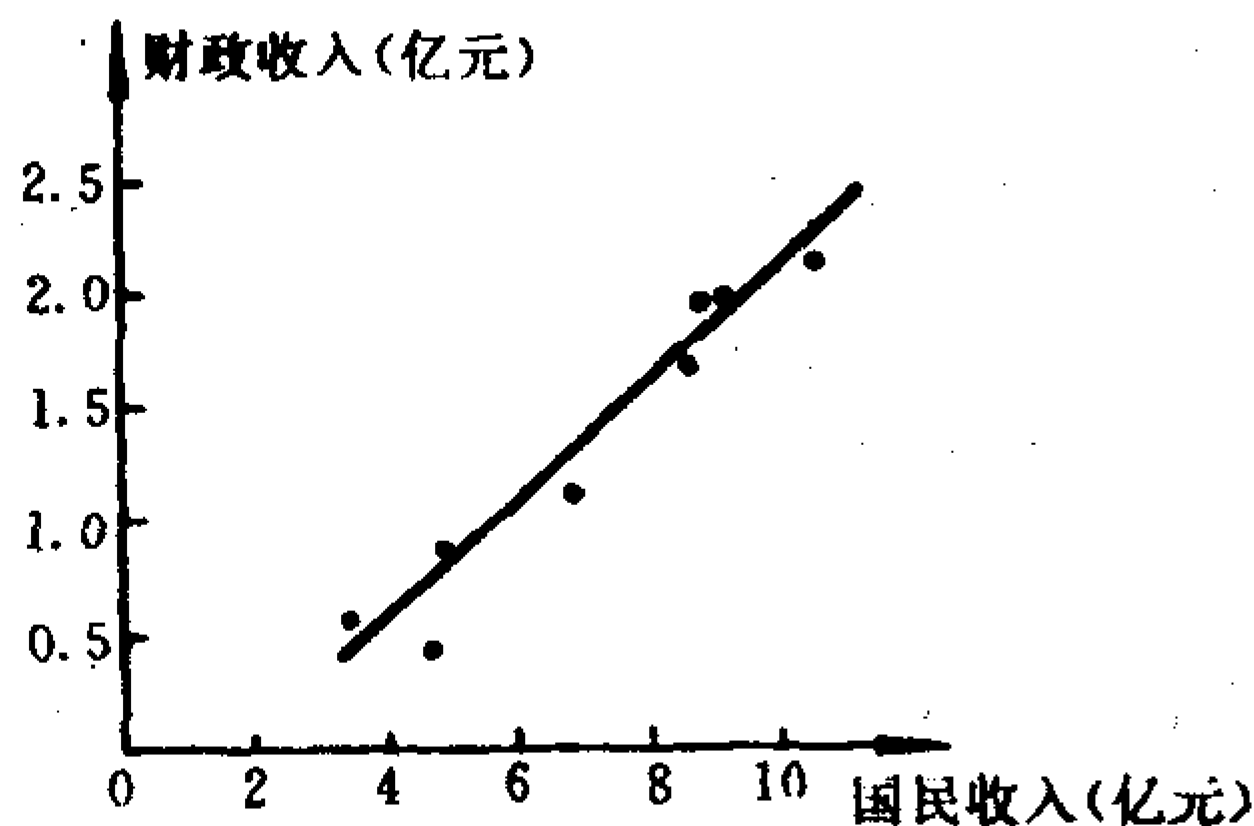


图 3.18

$$\Sigma x_i = 4.6 + 3.5 + \cdots + 9.5 = 55.2,$$

$$\Sigma y_i = 0.4 + 0.5 + \cdots + 2.2 = 10.2。$$

同理，有

$$\Sigma x_i^2 = 418.72, \Sigma x_i y_i = 81.43, \Sigma y_i^2 = 16.36。$$

$$\bar{x} = 55.2/8 = 6.9, \bar{y} = 10.2/8 = 1.275,$$

$$L_{xx} = \Sigma x_i^2 - \frac{1}{n}(\Sigma x_i)^2 = \Sigma x_i^2 - n(\bar{x})^2$$

$$= 418.72 - 8 \times (6.9)^2 = 37.84,$$

$$L_{xy} = \Sigma x_i y_i - \frac{1}{n}(\Sigma x_i)(\Sigma y_i) = \Sigma x_i y_i - n\bar{x}\,\bar{y}$$

$$= 81.43 - 8 \times 6.9 \times 1.275 = 11.05,$$

$$b = L_{xy}/L_{xx} = 11.05/37.84 = 0.292,$$

$$a = \bar{y} - b\bar{x} = 1.275 - 0.292 \times 6.9 = -0.740。$$

故该市国民收入与财政收入的回归直线为：

$$\hat{y} = -0.740 + 0.292x。$$

对求得的回归直线进行显著性检验，计算 r 值：

$$L_{yy} = \Sigma y_i^2 - \frac{1}{n}(\Sigma y_i)^2 = \Sigma y_i^2 - n(\bar{y})^2$$

$$= 16.36 - 8 \times 1.275^2 = 3.355。$$

$$r = L_{xy}/\sqrt{L_{xx}L_{yy}} = 11.05/\sqrt{37.84 \times 3.355}$$

$$= 0.98,$$

由相关系数临界值表查得

$$r_{0.01}(8-2) = 0.834。$$

因 $|r| > r_{0.01}$，故财政收入 y 与国民收入 x 的线性关系是高度显著的。

利用回归方程可以用自变量 x 对因变量 y 进行预测或控制，这里不再赘述。

2. 马尔可夫预测

马尔可夫(A. A. Markov) 链是一种特殊的概率模式，它在商业、经济学以及其他社会科学中都有着极为广泛的应用。

在一系列试验中，出现可列个两两互斥的事件 $E_i(i = 1,2,\cdots,n)$，且一次试验只出现其中的一个就称为状态。从一种状态变化到另一种状态称为状态转移。马尔可夫过程的特征是：若已知现在所处的状态，则将来的转移情况与过去所处的状态无关。即是说，已知系统现在所处的状态就能估计系统将来转移到各种状态的可能性，系统以前曾经处于什么状态对预测将来已不起作用。对确

定性现象，这一原则是明确的。例如，在质点的匀速直线运动中，若已知质点在 t_0 时所处的位置是 s_0，则有一般关系式

$$s = s_0 + vt$$

现在若已知在时刻 $t_1 > t_0$ 质点的位置是 s_1，则在将来的某一时刻 $t_2 > t_1$，质点所处的位置，可以预测为

$$s_2 = s_1 + v(t_2 - t_1)。$$

在这一预测式中，并没有出现"t_0 时刻质点曾经处于 s_0"的信息。亦即在已知现在 $s|_{t=t_1} = s_1$ 的情况下，将来的情况

$$s|_{t=t_2} = s + (v_2 - v_1),$$

$t_2 > t_1$ 与过去的情况 $s|_{t=t_0} = s_0$ 已没有直接的关系。

对于随机现象中的马尔可夫过程，预测其以后的状态，其思想方法是：

设当系统在 $k = 0$ 时的初始状态 $S^{(0)}$ 为已知时，经过 k 次转移后处在状态 i 的概率为 $S_i^{(k)}$，且有

$$\sum_{i=1}^{N} S_i^{(k)} = 1,$$

其中 N 为系统中互不相容的状态数。则一次转移后的状态为

$S^{(1)} = S^{(0)}P$，二次转移后为

$$S^{(2)} = S^{(1)} \cdot P = S^{(0)} \cdot P^2,$$

k 次转移后为

$$S^{(k)} = S^{(0)}P^k,$$

其中 P 为转移矩阵，即

$$P=\begin{bmatrix} P_{11} & P_{12} & \cdots\cdots & P_{1N} \\ P_{21} & P_{22} & \cdots\cdots & P_{2N} \\ \cdots\cdots & & & \\ P_{N1} & P_{N2} & \cdots\cdots & P_{NN} \end{bmatrix}。$$

转移矩阵是一个方阵，它的每一行是一个 $1\times N$ 阶的行矩阵 $[P_{i1},P_{i2},\cdots,P_{iN}]$ $(i=1,2,\cdots,N)$，其元素为非负，各元素之和为 1，这样的行矩阵通常称为 $1\times N$ 阶概率矩阵。这样，预测 k 次转移后的状态情况就化为计算 $S^{(k)}=S^{(0)}P^k$ 的问题。

例如，假定东南亚各国行销中国、日本和中国香港地区产的三种味精，现在要预测三个月后东南亚三种味精市场占有率的变动情况，以确定其对策。从国际市场调查知东南亚各国目前购买中国、日本、香港地区的三种味精的百分比分别为 40%、30%、30%。又知上月买中国味精的顾客本月仍有 40% 买中国味精，转买日本和香港地区味精的各占 30%；上月买日本味精的顾客本月仍有 30% 买日本味精，而有 60% 和 10% 分别转买中国和香港地区味精；上月买香港地区味精本月仍有 30% 买香港地区的味精，而有 60% 和 10% 分别转买中国和日本味精。由题意知，初始状态为

$$S^{(0)}=[0.4,0.3,0.3],$$

转移矩阵为

$$P=\begin{bmatrix} 0.4 & 0.3 & 0.3 \\ 0.6 & 0.3 & 0.1 \\ 0.6 & 0.1 & 0.3 \end{bmatrix},$$

$$S^{(3)}=S^{(0)}P^3=[0.4,0.3,0.3]\begin{bmatrix} 0.4 & 0.3 & 0.3 \\ 0.6 & 0.3 & 0.1 \\ 0.6 & 0.1 & 0.3 \end{bmatrix}^3$$

$$= [0.5008, 0.2496, 0.2496]。$$

即三月后，东南亚市场中国、日本、香港地区对三种味精市场占有率的预测值分别为 50%、25%、25%。

若要预测较长时间后，马尔可夫过程趋于稳定状态的情况，其思想方法是：

由于趋于稳定状态时 $S^{(n)} = S^{(n-1)}$，且

$$\sum_{i=1}^{N} S_i^{(n)} = 1。 \quad ①$$

又 $S^{(n)} = S^{(n-1)}P$，则

$$S^{(n)} = S^{(n)}P。 \quad ②$$

改写为矩阵式为：

$$[S_1^{(n)}, S_2^{(n)}, \cdots, S_N^{(n)}] = [S_1^{(n)}, S_2^{(n)}, \cdots, S_N^{(n)}] \cdot \begin{bmatrix} P_{11} \ P_{12} \cdots\cdots P_{1N} \\ P_{21} \ P_{22} \cdots\cdots P_{2n} \\ \cdots\cdots \\ P_{N1} \ P_{N2} \cdots P_{NN} \end{bmatrix}。$$

① 与 ② 联立得方程组

$$\begin{cases} S_1^{(n)} = S_1^{(n)}P_{11} + S_2^{(n)}P_{21} + \cdots + S_N^{(n)}P_{N1}, \\ S_2^{(n)} = S_1^{(n)}P_{12} + S_2^{(n)}P_{22} + \cdots + S_N^{(n)}P_{N2}, \\ \cdots\cdots \\ S_N^{(n)} = S_1^{(n)}P_{1N} + S_2^{(n)}P_{2N} + \cdots + S_n^{(n)}P_{NN}, \\ S_1^{(n)} + S_2^{(n)} + \cdots + S_N^{(n)} = 1。 \end{cases}$$

由于此方程组中前 N 个方程中有一个是不独立的，若消去第 N 个方程，并将其余 N 个方程整理得

$$\begin{cases} (P_{11} - 1)S_1^{(n)} + P_{21}S_2^{(n)} + \cdots + P_{N1}S_N^{(n)} = 0, \\ P_{12}S_1^{(n)} + (P_{22} - 1)S_2^{(n)} + \cdots + P_{N2}S_N^{(n)} = 0, \\ \cdots\cdots \\ S_1^{(n)} + S_2^{(n)} + \cdots + S_N^{(n)} = 1。 \end{cases} \quad ③$$

亦即

$$\begin{bmatrix} (P_{11}-1) & P_{21} & \cdots & P_{N1} \\ P_{12} & (P_{22}-1) & \cdots & P_{N2} \\ \cdots\cdots & & & \\ 1 & 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} S_1^{(n)} \\ S_2^{(n)} \\ \vdots \\ S_N^{(n)} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}。④$$

令

$$P^* = \begin{bmatrix} (P_{11}-1) & P_{21} & \cdots & P_{N1} \\ P_{12} & (P_{22}-1) & \cdots & P_{N2} \\ \cdots\cdots & & & \\ 1 & 1 & \cdots & 1 \end{bmatrix}, B = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix},$$

则 ④ 为 $P^* S^{(n)} = B$。若$(P^*)^{-1}$存在，便有$S^{(n)} = (P^*)^{-1}B$。这就是稳定状态概率，它与初始状态无关。

例如，上例中，

$$P = \begin{bmatrix} 0.4 & 0.3 & 0.3 \\ 0.6 & 0.3 & 0.1 \\ 0.6 & 0.1 & 0.3 \end{bmatrix},$$

$$P^* = \begin{bmatrix} -0.6 & 0.6 & 0.6 \\ 0.3 & -0.7 & 0.1 \\ 1 & 1 & 1 \end{bmatrix},$$

故

$$S^{(n)} = \begin{bmatrix} -0.6 & 0.6 & 0.6 \\ 0.3 & -0.7 & 0.1 \\ 1 & 1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

$$= \frac{1}{0.96} \begin{bmatrix} A_{11} & A_{21} & 0.48 \\ A_{12} & A_{22} & 0.24 \\ A_{13} & A_{23} & 0.24 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.50 \\ 0.25 \\ 0.25 \end{bmatrix}。$$

即中国、日本、香港地区三种味精经较长时间后在市场的

占有率趋于稳定状态，分别为 50%、25%、25%。这个稳定状态，与初始状态无关，即使初始状态为其他数值，如 $S^{(0)}=[0.7,0.2,0.1]$ 等，经较长时间后仍可达到[0.50，0.25，0.25]的稳定状态。

本例表明，决定产品在市场中占有率的情况，从长远来看，不能看初始状态，而主要取决于转移矩阵。因此，一种产品只有不断提高产品质量，降低成本，改善流通和经营方式，才能进一步巩固和扩大市场。若眼光只注意原来的初始状态，不注意顾客对产品的转移情况，则在激烈的市场竞争中必将会吃败仗。

"已知现在，则将来与过去无关"的马氏原则是一个十分重要的原则。在自然界中，有很多现象是符合或近似符合这个原则的。现在，马氏预测的思想已广泛应用于生物学、医学、经济等部门。作为应用，我们考察下面的事例。

例 2　股票市场。

对某防污技术公司的股票价格变化进行研究后发现：

如果当天的股票上涨，则第二天其价格上涨、不变、下跌的概率分别是 $\frac{5}{8}$、$\frac{1}{8}$、$\frac{2}{8}$；如果当天的股票价格不变，则第二天其价格上涨、不变、下跌的概率分别是 $\frac{2}{8}$、$\frac{1}{8}$、$\frac{5}{8}$；如果当天的股票下跌，则第二天其价格上涨、不变、下跌的概率分别是 $\frac{3}{4}$、$\frac{1}{8}$、$\frac{1}{8}$。已知某一天的状态矢为 $\left[\frac{5}{8},\frac{1}{8},\frac{2}{8}\right]$，问第二天该股票的市场价格上涨、不变和下跌的概率分别为多少？该公司股票价格上涨、不变和下

跌的稳态概率是什么?

据题意,$S^{(0)}=\left[\frac{5}{8},\frac{1}{8},\frac{2}{8}\right]$,转移矩阵

$$P=\begin{bmatrix}\frac{5}{8} & \frac{1}{8} & \frac{2}{8}\\ \frac{2}{8} & \frac{1}{8} & \frac{5}{8}\\ \frac{6}{8} & \frac{1}{8} & \frac{1}{8}\end{bmatrix}。$$

则

$$S^{(1)}=S^{(0)}P=\left[\frac{5}{8},\frac{1}{8},\frac{2}{8}\right]\begin{bmatrix}\frac{5}{8} & \frac{1}{8} & \frac{2}{8}\\ \frac{2}{8} & \frac{1}{8} & \frac{5}{8}\\ \frac{6}{8} & \frac{1}{8} & \frac{1}{8}\end{bmatrix}$$

$$=\left[\frac{39}{64},\frac{1}{8},\frac{17}{64}\right]。$$

故第二天该公司的股票价格上涨、不变、下跌的概率分别是$\frac{39}{64}$,$\frac{1}{8}$和$\frac{17}{64}$。

求稳态概率,只要按公式$S^{(n)}=(p^{*})^{-1}B$求即可。现

$$P^{*}=\begin{bmatrix}-\frac{3}{8} & \frac{2}{8} & \frac{6}{8}\\ \frac{1}{8} & -\frac{7}{8} & \frac{1}{8}\\ 1 & 1 & 1\end{bmatrix},$$

故

$$S^{(n)}=(P^*)^{-1}B=\frac{8}{9}\begin{bmatrix}A_{11} & A_{21} & \frac{11}{16}\\ A_{12} & A_{22} & \frac{9}{64}\\ A_{13} & A_{23} & \frac{19}{64}\end{bmatrix}\begin{bmatrix}0\\0\\1\end{bmatrix}$$

$$=\left[\frac{11}{18},\frac{1}{8},\frac{19}{72}\right]。$$

即该公司股票价格上涨、不变与下跌的长期稳态概率分别是$\frac{11}{18}$、$\frac{1}{8}$与$\frac{19}{72}$。它表明，其股票在长期变化过程中，上涨的情况占$\frac{11}{18}\approx 61.1\%$，不变的情况占$\frac{1}{8}=12.5\%$，下跌的占$\frac{19}{72}\approx 26.4\%$。

本题的稳态概率不用公式，还可直接从方程组求得。方法是：

设$[x,y,z]$是所求的稳态概率，则有

$$\begin{cases}[x,y,z]=[x,y,z]P\\ x+y+z=1\end{cases}\Leftrightarrow\begin{cases}3x-2y-6z=0\\ -x+7y-z=0\\ 2x+5y-7z=0\\ x+y+z=1\end{cases}$$

$$\Leftrightarrow\begin{cases}3x-2y-6z=0\\ -x+7y-z=0\\ x+y+z=1\end{cases}\Leftrightarrow\begin{cases}x=\frac{11}{18}\\ y=\frac{1}{8}\\ z=\frac{19}{72}。\end{cases}$$

例3 城乡人口的迁移。

设一城市和农村的总人口为a，又记在城市为状态A，在农村为状态B。已知$A\to A$与$A\to B$的概率分别为

0.98 和 0.02；$B \to A$ 与 $B \to B$ 的概率分别为 0.03 和 0.97。求其人口稳定状态的情况。

据条件件知，其转移矩阵为

$$P = \begin{pmatrix} 0.98 & 0.02 \\ 0.03 & 0.97 \end{pmatrix}, \text{则 } P^* = \begin{pmatrix} -0.02 & 0.03 \\ 1 & 1 \end{pmatrix}。$$

$$S^{(n)} = (P^*)^{-1}B = \frac{1}{-0.05}\begin{pmatrix} A_{11} & -0.03 \\ A_{12} & -0.02 \end{pmatrix}\begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

$$= \begin{pmatrix} 0.6 \\ 0.4 \end{pmatrix}。$$

由已知条件知城市和农村的总人口为 a，故经较长时间后，城市和农村的人口分别趋向于 $0.6a$ 和 $0.4a$。

3.9 数学归纳原理与数学归纳法

1. 自然数公理与数学归纳原理

数学归纳法原理是建筑在自然数公理系统的基础上的。这个公理系统由意大利数学家皮亚诺于 1891 年创立，故又称皮亚诺公理。它以 1、后继为原始概念，包含有以下五条公理：

1° 设 N 为非空集合，$1 \in N$；

2° 任给 $a \in N$，在 N 中存在 a 的后继 a'；

3° 若 $a' = b'$，则 $a = b$。即若后继相等，其前驱亦相等；

4° 任给 $a \in N$，有 $a' \neq 1$，即 1 没有前驱；

5° 若 $M \subseteq N$，且 $1 \in M$，又若 $n \in M$，就有 $n' \in M$，则 $M = N$。

满足公理 1°—5° 的集合 N 称为自然数集合*，其中公理 5° 称为归纳公理。对自然数集合而言，它是不可缺少的，若缺少 5°，不足以刻画自然数。事实上，设有

$$1,2,3,\cdots,\omega,\omega+1,\omega+2,\cdots \qquad (*)$$

其中 ω 按康托的观点，是 $1,2,3,\cdots,n,\cdots$ 中在 n 无限增大时趋向的目标，亦即为极限序数，于是（*）满足公理 1°—4°。这说明仅满足 1°—4° 的数虽然包括了我们所需的自然数，但它还可以包括更多的数。但若加上公理 5°，由于 ω 不是任何数的后继，而 $\omega,\omega+1,\omega+2,\cdots$ 只是从 ω 开始通过后继得到的数，故满足 1°—5° 的数不能包括 $\omega,\omega+1,\omega+2,\cdots$ 只能包括从 1 开始通过后继得到的数 $1,2,3,\cdots$ 即我们所需的自然数。故满足皮亚诺公理的数恰是我们所需的自然数，亦即皮亚诺公理定义了自然数。由皮亚诺自然数公理中的归纳公理，可得如下数学归纳法合理性定理：

设 $T(n)$ 表示叙述中含自然数 n 的某一命题，若

1° $T(1)$ 真；

2° 任给 $n_0\in N$，若 $T(n_0)$ 真 $\rightarrow T(n_0')$ 真，则对一切 n，$T(n)$ 真。

证明：若记

$M=\{k\in N|T(k)\text{ 真}\}$，

亦即 M 是使命题 $T(k)$ 为真的自然数 k 的集合。由于

---

* 从集合论观点，既然集合论是整个经典数学的基础，那么任何数学理论应渊源于集合论，自然数系统及其算术理论，当亦不应例外，就是说我们应当用构造性的方法从集合的概念出发，去把自然数的概念定义出来。如何从集合概念出发用构造性方法去建立自然数系统，最终把整个算术理论嵌入集合论，有兴趣的读者可参阅文献[31]第四章第一节的有关内容。

$T(1)$ 真，故 $1 \in M$；又由 $T(n_0)$ 真 $\rightarrow T(n_0')$ 真，则知 $n_0 \in M \rightarrow n_0' \in M$，根据皮亚诺归纳公理必有 $M = N$，这意味着，对任意自然数 n，$T(n)$ 为真。

关于数学归纳原理，还可用如下良序公理 5* 来代替等价的皮亚诺归纳公理 5°。

5* 良序公理：由自然数全体构成的有序集 N 中任何非空有序子集 S 必有最小元。即 $\exists\ s_0 \in S$，对 $\forall\ s \in S$，$s_0 \leqslant s$。

其中，良序集是指这样的一个有序集，当且仅当它的每一个非空子集有一个最小元 s_0。

例如，整数集 Z，按数目大小关系"$\leqslant$"是有序集，但不是良序集。因为其非正整数集

$$A = \{x \mid x \in Z, x \leqslant 0\}$$

非空，但没有最小元，故 $(Z, \leqslant)$ 不是良序集。但若规定如下的顺序：0 小于任何整数；凡绝对值相同的整数中，规定负数小于正数，而绝对值不同的整数，规定绝对值小的整数较小。对这样的顺序，Z 是良序集。一般地，对于任意集合，总可以对它规定适当顺序，使之成为良序集，亦即任意集合 A 均可良序化。

由良序公理 5*，可得如下的数学归纳法第二形式(即第二数学归纳法)合理性定理：

设有某一含有自然数 n 陈述的定理 T，若 $n = 1$ 定理 T 真，又对 $n < n_0$ 定理真能推得 $n_0(n_0 > 1)$ 定理亦真，则对任何 n 定理 T 皆真。

证明：构造集合

$$M = \{n \in N \mid T(n) \text{ 不成立}\}。$$

即 M 是由使定理 T 不成立的自然数所组成的集合。现假设数学归纳法第二形式不成立，即不是对任何 n 定理 T

皆真，则M必为非空。根据良序公理，非空集M中必有最小数，设为n_0，$n_0 \in M$。

由于对于数1，定理T真，故$1 \notin M$，从而$n_0 \neq 1$。又在满足自然数公理1°—5°（5*与5°等价）的条件下知1为最小的自然数，必有$n_0 > 1$。由于n_0是M中的最小数，则小于n_0的自然数便不属于M，即$n < n_0$时定理T为真。由定理条件，对n_0，定理T亦真，即$n_0 \notin M$，这与前面结论$n_0 \in M$矛盾，故M不能为非空，即$M = \varnothing$（空集），从而第二数学归纳法为真。

由于任意集合都可以良序化，即可以适当规定顺序，使之成为良序集，故有如下的超限归纳原理：

设$(W, \leqslant)$是一个良序集，$P(x)$是涉及W中任意x的一个命题，如果

1° 对于W中最小元W_1，$P(W_1)$成立；

2° 假定对任意$x < a$，$P(x)$成立，可证$P(a)$成立；

那么，对W中任意元x，$P(x)$均成立。

超限归纳法对任意集合都能适用，而普通数学归纳法只适用于自然数集。因每一自然数都是有限数，故普通数学归纳法也叫有限归纳法，以与超限归纳法相区别。

2. 数学归纳法及其使用技巧

由前面可知，数学归纳法是依据数学归纳原理进行证明的重要方法。其思维实质是：

若一个含自然数n的陈述为P，记$P(1)$为A，$\forall n(P(n) \to P(n+1))$为$B$，$\forall n, P(n)$表示为$C$，则数学归纳法的推理过程为，从$A$、$B$、$A \land B \to C$。这是一种演绎性推理。故数学归纳法的作用在于它将无限的归纳推理变为有限的演绎推理。正是由于这样的原因，应用数学归纳法作出的结论是可靠的。

数学归纳法的论证步骤有二，一是归纳奠基，二是递推。在使用数学归纳法的过程中，由于不理解其思维原理，常产生各种形式的错误。常见的有：

(1) 形式套用归纳步骤。

例如，求证：所有人的年龄是一样的。

用数学归纳法证明如下：

1° 当 $n=1$ 时，命题显然成立。

2° 假设命题对 n 成立，即对任何 n 个人，其年龄一样，证对 $n+1$ 个人命题亦成立。

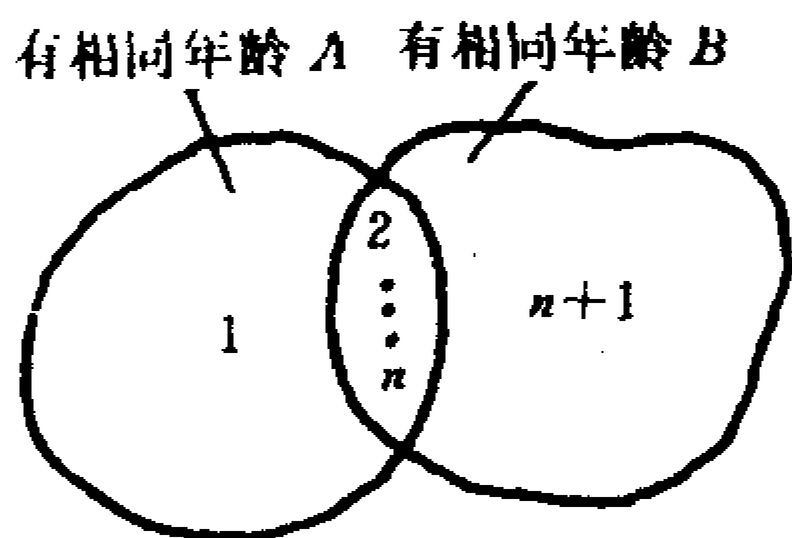


图 3.19

设将这 $n+1$ 个人编号，记为 $1,2,\cdots,n,n+1$。于是 $1,2,\cdots,n-1,n$ 共 n 个。由归纳假定，有相同年龄 A；同时 $2,3,\cdots,n,n+1$ 共 n 个也有相同年龄 B，于是，$2,3,\cdots,n$ 和 1 同为 A，又和 $n+1$ 同为 B。这样，$2,3,\cdots,n$ 的年龄既是 A，又是 B，于是 $A=B$。即假设由 $P(n)$ 成立，推到了 $P(n+1)$ 亦成立。故所有人的年龄是一样的。

结论是荒谬的。错误在于 $n=2$ 时，命题不成立，导致推断中断。

(2) 不用归纳假设。

例如，证明：$\sqrt{n^2+n}<n+1$。$(n\in N)$

错误证法是：

1° 当 $n=1$ 时，左边 $=\sqrt{1^2+1}=\sqrt{2}<2=1+1$ $=$ 右边，故不等式成立。

2° 假设 $n=k$ 时不等式成立，即

$\sqrt{k^2+k}<k+1$，

那么

$$\sqrt{(k+1)^2+(k+1)}=\sqrt{k^2+3k+2}$$

$$<\sqrt{k^2+4k+4}=\sqrt{(k+2)^2}=k+2=(k+1)+1。$$

这说明，当 $n=k+1$ 时，不等式亦成立。

故由 1°、2° 知对任何自然数 n 不等式都成立。

如上推导 $n=k+1$ 时不等式成立未用到归纳假设，而是采用了被开方数"放大"的技巧，这不是数学归纳证法。

正确的证法，2° 应为：

设 $n=k$ 时不等式成立，即

$$\sqrt{k^2+k}<k+1\Leftrightarrow k^2+k<(k+1)^2。$$

则

$$\sqrt{(k+1)^2+(k+1)}=\sqrt{(k^2+k)+(2k+2)}$$

$$<\sqrt{(k+1)^2+(2k+2)}=\sqrt{k^2+4k+3}$$

$$<\sqrt{k^2+4k+4}=\sqrt{(k+2)^2}=k+2=(k+1)+1。$$

即当 $n=k+1$ 时不等式亦成立。

(3) 混淆归纳变量。

例如，求证：$C_n^k=\dfrac{n!}{k!(n-k)!}$。

对数学归纳法的第二步，有人这样证明：

假设当 $n=k$ 时，命题为真，即

$$C_k^k=\frac{k!}{k!(k-k)!}=1。$$

当 $n=k+1$ 时，一方面 $C_{k+1}^k=k+1$，另一方面

$$C_{k+1}^k=\frac{(k+1)!}{k![(k+1)-k]!}=k+1。$$

故对一切自然数 n，命题是正确的。

上述在证明 $C_{k+1}^k=k+1$ 时丢开了归纳假设，实际

充其量只仅仅在进行特殊值的验证。更严重的是，证明混淆了归纳变量。事实上，应该把 n 看作常量，对变量 k 进行归纳。因为否则，当 $k>1$ 时，就无法讨论 $n=1$ 时的情况了。

正确的证法应是：

1° 当 $k=1$ 时，

$$C_n^1=n,\quad \frac{n!}{k!(n-k)!}=\frac{n!}{1!(n-1)!}=n,$$

命题正确。

2° 假设当 $k=r$ 时命题为真，即

$$C_n^r=\frac{n!}{r!(n-r)!}=\frac{n(n-1)\cdots(n-r+1)}{r!}.$$

则当 $k=r+1$ 时，

$$\begin{aligned}C_n^{r+1}&=\frac{n!}{(r+1)!(n-r-1)!}\\&=\frac{n(n-1)\cdots(n-r+1)(n-r)}{1\cdot 2\cdots r(r+1)}\\&=C_n^r\frac{n-r}{r+1}。\end{aligned}$$

因为左端表示从 n 个元素中每次取出 $r+1$ 个元素的组合种数，它等于把集合里 n 个元素分成元素为 r 个及 $n-r$ 个的两个子集，然后先从 n 个元素中每次取出 r 个元素的组合种数，再从 $n-r$ 个元素中任选一个元素添加进去，这样共有 $n-r$ 个搭配，但重复计算了 $(r+1)$ 次，于是，

$$C_n^{r+1}=C_n^r\frac{n-r}{r+1},$$

故当 $k=r+1$ 时结论为真。

由此断定对一切自然数 k 和 $n(k\leqslant n)$ 命题是正确的。

在应用数学归纳法证题时，存在若干思维技巧，主要有：

(1) 灵活选取归纳“起点”与“跨度”。

通常的数学归纳法，常以验证 $n=1$ 时的命题入手，这就叫做以 $n=1$ 作为起点；归纳的第二步也常为由 $n=k$ 成立推出 $n=k+1$ 成立，这叫做以跨度 1 前进。这都是一些最基本的程式，但起点和跨度应因势而异，灵活选取。

例如，证明：用 3 分和 5 分邮票可以组成 8 分以上的任何邮资。

证明：1° 因为 $8=3+5$，$9=3+3+3$，故用 3 分、5 分邮票可组成 8 分，9 分邮资。

2° 设当 $k\geqslant 9$ 时结论成立，要证结论对 $k+1$ 亦成立。

若 k 分邮票中有一张 5 分邮票，则用 2 张 3 分邮票换取这张 5 分邮票，就可凑得 $k+1$ 分邮资。若 k 分邮票中无 5 分邮票，则它至少有 3 张 3 分邮票(因 $k\geqslant 9$)，那么用 2 张 5 分邮票换去 3 张 3 分邮票，又可凑得 $k+1$ 分邮资。故结论对 $k+1$ 亦成立。

由 1°、2° 便知任何 8 分以上的邮资，均可用 3 分、5 分两种邮票组成。

本例选取 $n=8$、9 作为归纳起点，利于第二步的证明，即便于从 $n=k$ 时结论成立推出 $n=k+1$ 时结论成立。

(2) 选择合适的假设方式。

归纳假设也需因势而异，并非一定拘泥于 $n=k$ 的程式，可以是 $n\leqslant k$，$n<k$ 或 $n=k+1$ 等。

例如，设有颗数均为 n 的两堆棋子，两人进行如下的比赛：

每人在且只在其中一堆里任取几颗，规定取完最后一颗者胜利。求证：后取者必胜。

证明：1° $n = 1$，即每堆仅有一颗棋子，显然后取者必胜。

2° 设 $n < k$ 时结论为真，要证 $n = k$ 时结论亦真，即是证假定每堆棋子数少于 k 时，后取者胜，能推得每堆棋子数为 k 时，也是后取者胜。

因为先取者在某堆中取了 l 颗（$l \leqslant k$），则后取者可在另一堆也取 l 颗，这时每堆还有 $k - l$ 颗，因 $k - l < k$，由归纳假设知后取者胜利，这说明在 $n = k$ 时结论亦成立。

由 1°、2° 知，结论对任意自然数 n 都成立。

本例选取 $n < k$ 时结论为真作为归纳假设，即用第二数学归纳法证明，较利于问题的证明。

(3) 改变归纳途径。

归纳的一般途径是只进不退，即常在命题对较小的 n 成立的假设下去推演命题对较大的 n 成立。但在这种推进受阻时，可以改变为“有进有退，进退结合”的策略。

例如，设对应法则 f 对一切自然数 n 有意义，且

(1) $f(n)$ 是整数；

(2) $f(2) = 2$；

(3) 对一切自然数 m 和 n，$f(mn) = f(m) \cdot f(n)$；

(4) 当 $m > n$ 时，$f(m) > f(n)$。

试证：$f(n) = n$。

证明：1° $2 = f(2) = f(1 \cdot 2) = f(1) \cdot f(2) = 2f(1)$，

即 $f(1) = 1$，所以命题当 $n = 1$ 时成立。

2° 假设对 $n \leqslant k$ 命题成立，即有 $f(n) = n$，要证

$f(k+1)=k+1$。

注意到要应用关系式

$f(mn)=f(m)\cdot f(n)$,

故需对 $k+1$ 的奇偶性进行讨论。

① 若 $k+1$ 为偶数,则令 $k+1=2j$,其中 j 为正整数,且 $j\leqslant k$,于是有 $f(j)=j$,

$f(k+1)=f(2j)=f(2)\cdot f(j)=2j=k+1$。

故当 $k+1$ 为偶数时,命题成立。

② 若 $k+1$ 为奇数,则令

$k+1=2j+1(j<k)$。

现先考察 $f(k+2)$,因 $k+2$ 为偶数,故

$k+2=(2j+1)+1=2(j+1)\ (j+1\leqslant k)$,

$$\begin{aligned}f(k+2)&=f(2\cdot(j+1))=f(2)\cdot f(j+1)\\&=2(j+1)=k+2。\end{aligned}$$

再考察 $f(k+1)$。由(4)、(1),知

$f(k)<f(k+1)<f(k+2)$,

且 $f(k+1)$ 为整数,故由 $f(k)=k$ 及

$f(k+2)=k+2$

可知 $f(k+1)=k+1$,所以当 $k+1$ 时命题成立。

由 1°、2° 知,对一切自然数 n,$f(n)=n$。

本例中,对 $k+1$ 为奇数时的归纳途径是:

$n\leqslant k\Rightarrow n=k+2\Rightarrow n=k+1$,

即用留下 $k+1$ 的空位后再补遗的方法进行归纳。

3.10　同构原理与同构方法

1. 同构与同构映射

两集合 M 与 N,当它们建立了元素间的一一对应

后，集合 M 间元素的关系、运算等未必类同于集合 N 的元素的关系和运算。但也有许多特殊情况，即两集合 M 与 N 间的一一对应，使集合 M 间元素的关系或运算或满足的公理类同于集合 N 间元素的关系、运算和满足的公理等。这时，我们就将这样的两个集合看作是形式相异而本质相同的事物，并称它们为同构的。即同构所揭示的是表面上不同的对象和关系的体系在结构上的深刻类似性。

例如，$A=\{15,16,10,13\}$ 是 4 人年龄的集合，$B=\{40,42,44,46\}$ 是制服号码的集合。设有对应法则：

$$f_1: 10\to 40, 13\to 42, 15\to 44, 16\to 46。$$

f_1^{-1} 是使箭头相反的对应。这时，f_1 和 f_1^{-1} 中每一个都是保持“小于”或“大于”关系的一一对应，这种一一对应就称为保持小于或大于关系的同构。故 A、B 就小于关系或大于关系而言是同构的。又如，直线上点的次序 $\prec$（…先于…）与实数的小于关系也是同构的。

在数学中，尤为重要的是保持运算的同构。

设集合 M 与 $\overline{M}$ 分别有代数运算 $\circ$ 与 $\bar{\circ}$，若 M 与 $\overline{M}$ 的一一对应 f，使

$$a\to\bar{a}, b\to\bar{b}, a\circ b\to\bar{a}\,\bar{\circ}\,\bar{b}。$$

其中 a、$b\in M$，$\bar{a}$、$\bar{b}\in\overline{M}$，则说一一对应 f 是 M 到 $\overline{M}$ 的同构对应。

例如，$f_2: x\to 2x, x\in N$。

$$\begin{aligned} f_2(x_1+x_2) &= 2(x_1+x_2)=2x_1+2x_2 \\ &= f_2(x_1)+f_2(x_2)。\end{aligned}$$

即它是保持加法运算的同构，且此对应中，又保持了小于关系。若以 B 表示正偶数集合，则此同构可记作

$$[N, <, +] \underset{f_2^{-1}}{\overset{f_2}{\rightleftarrows}} [B, <, +]。$$

又如，$f_3: x \to 2^x, x \in R$。这是实数集 R 到正实数集 R^+ 的一一映射。

$f_3(x_1 + x_2) = 2^{x_1 + x_2} = 2^{x_1} \cdot 2^{x_2} = f_3(x_1) \cdot f_3(x_2)$，$x_1$、$x_2 \in R$。

$f_3^{-1}: y \to \log_2 y, y \in R^+$。

$f_3^{-1}(y_1 \cdot y_2) = \log_2(y_1 \cdot y_2)$

$= \log_2 y_1 + \log_2 y_2$

$= f_3^{-1}(y_1) + f_3^{-1}(y_2)$，$y_1$、$y_2 \in R^+$。

故

$$[R, <, +] \underset{f_3^{-1}}{\overset{f_3}{\rightleftarrows}} [R^+, <, \cdot]。$$

亦即 f_3 是 R 到 R^+，f_3^{-1} 是 R^+ 到 R 的同构映射，且都保持小于关系。而 $f_4: x \to \left(\frac{1}{2}\right)^x$，$x \in R$，是 R 到 R^+ 的同构映射。但这时，映射把小于关系转换为大于关系，即

$$[R, <, +] \underset{f_4^{-1}}{\overset{f^4}{\rightleftarrows}} [R^+, >, \cdot]。$$

若集合 M 对运算 $\circ$ 构成群，又集合 $\overline{M}$ 对运算 $\bar{\circ}$ 构成群，且映射 f 使

$$a \circ b \xrightarrow{f} \bar{a}\,\bar{\circ}\,\bar{b}(a、b \in M, \bar{a}、\bar{b} \in \overline{M})，$$

则称群$(M, \circ)$与$(\overline{M}, \bar{\circ})$是同构的，而 f 是群 M 到 $\overline{M}$ 的同构映射。

例如，$[R^+, \cdot]$、$[R, +]$ 分别构成一个群。前者是正实数对于数目乘法构成的群，后者是实数对于数目加法构成的群。于是，f_3、f_4 都是群$[R, +]$到群$[R^+, \cdot]$的同构映射，而 f_3^{-1}、f_4^{-1} 都是群$[R^+, \cdot]$到群$[R, +]$的同构映射。

2. 同构方法的功能及其应用

由于成同构的事物形异实同，故凡结构相同的数学对象不必一一加以研究，只需选择最简单的一种作为代表来加以研究，其余的只需根据对应法则，便可得出相应的结果。因而可收到事半功倍的效果。

(1) 命题代数与开关代数的同构。

集合代数、命题代数和开关代数都是布尔代数的模型，其关系如下表。

抽象的布尔代数	集合代数	命题代数	开关代数
集合 M 的元素	某集合的子集	二值命题	二个状态的开关
$\oplus$	并($\cup$)	析取($\vee$)	并联($\sqcup$)
$\odot$	交($\cap$)	合取($\wedge$)	串联($\sqcap$)
e_1	$\varnothing$(空集)	F	0(开)
e_2	I(全集)	T	1(闭)
x'	$\overline{A}$：集合 A 的补集	$\overline{X}$：命题 X 的否	$\overline{X}$：开关 X 的反相

由表中可知，抽象布尔代数的运算规则，对三个具体代数都适用。特别地，命题代数和开关代数都是二值布尔代数，它们是同构的。这种同构，使得专门建立开关代数的理论成为不必须，即它可以借助命题代数的工具来解决开关电路的分析、化简等问题。例如，假定要化简线路(图3.20(1))，即要做一个与它“等效的”而开关个数又较少的线路，由于给定线路对应于用命题代数的语言表示的下列式子：

$$(A \wedge B) \vee (\overline{A} \wedge C) \vee (B \wedge C), \qquad ①$$

借用命题代数语言，可将 ① 化为

$$(A \wedge B) \vee (\overline{A} \wedge C)。 \qquad ②$$

这是由于

$$B \wedge C = (A \vee \overline{A}) \wedge B \wedge C,$$

将它代入 ①，得

$$(A \wedge B) \vee (\overline{A} \wedge C) \vee (A \wedge B \wedge C) \vee (\overline{A} \wedge B \wedge C) = (A \wedge B) \vee (A \wedge B \wedge C) \vee (\overline{A} \wedge C) \vee (\overline{A} \wedge B \wedge C) = (A \wedge B) \vee (\overline{A} \wedge C),$$

即为 ②，对应于图 3.20(2) 的线路。

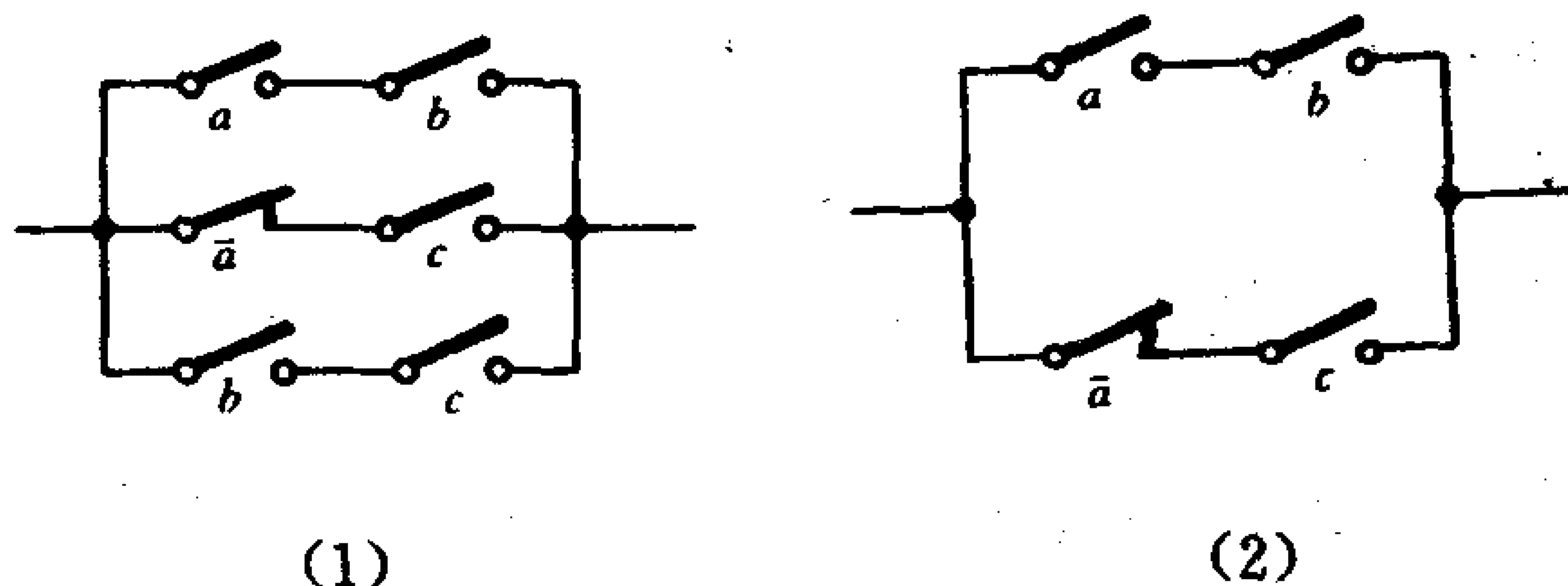


图 3.20

(2) 整除理论与集合代数的同构。

设有自然数 a 分解成质因数

$$a = a_1 \cdot a_2 \cdots a_n,$$

将各个质因数写下，且质因数相同的给以编号，这样就构成一个集合 $A = \{a_1, a_2, \cdots, a_n\}$。

例如，$a = 18 = 2 \times 3 \times 3$，对应的集合为 $A = \{2_1, 3_1, 3_2\}$，$b = 90 = 2 \times 3 \times 3 \times 5$，则 $B = \{2_1, 3_1, 3_2, 5_1\}$。这时，对应法则 f 使每个自然数 x 对应于它的质因数的集合 X，并把 x 整除 y 的关系 $x|y$ 转换成 X 包含于 Y 的关系 $X \subseteq Y$，它们是同构的。

因为任意自然数都可唯一地表示成它的质因数的乘积(不管因数的顺序)，所以在映射 f 下每个自然数 x 都对应唯一的集合 X，且不同的自然数 x 和 y 对应不同的集合 X 和 Y，即 f 是一一的。

同时,可以证明对于 N 中任意的 x 和 y,有

$x|y \Leftrightarrow X \subseteq Y$。

事实上,如果 $x|y$,那么 $y = xz$,把 x 和 z 分解质因数,就得到 y 的分解式。因为 y 可以唯一地表示成质因数的乘积,故在 y 的质因数中可以找到 x 的所有质因数,即 $X \subseteq Y$。

反过来,假定 $X \subseteq Y$,而且

$X = \{x_1, x_2, \cdots, x_k\}, Y = \{y_1, y_2, \cdots, y_n\}$。

那么所有 $x_i \in Y$。设 $x_1 = y_1, x_2 = y_2, \cdots, x_k = y_k$,即

$y = x_1 x_2 \cdots x_k y_{k+1} \cdots y_n$,

那么 $y = xz$,其中 $z = y_{k+1} \cdots y_n$,即 $x|y$,所以

$x|y \Leftrightarrow X \subseteq Y$。

在这样的同构基础上,整除性理论的一切概念、命题和证明都能翻译成集合代数的语言。比如,

$(d|x) \wedge (d|y) \Rightarrow (D \subseteq X) \cap (D \subseteq Y)$

$\Rightarrow D \subseteq X \cap Y$。

于是可得,"d 是 x 和 y 的最大公因数",可转换成"D 等于集合 X、Y 的交"。据此,还可证明:

1° 如果 a 整除 b,则 a 是 a 和 b 的最大公因数。

$a|b \Rightarrow A \subseteq B, A \subseteq B \Rightarrow A \cap B = A$,

由 $A \cap B = A \Rightarrow a$ 是 a、b 的最大公因数。

2° 如果 c 整除 a 和 b 的每一个,那么它整除它们的最大公因数。

$(c|a) \wedge (c|b) \Rightarrow (C \subseteq A) \cap (C \subseteq B) \Rightarrow C \subseteq A \cap B$

$\Rightarrow c$ 整除 a、b 的最大公因数。

(3) 复数的同构。

复数可以认为是形如 $a + bi$ 的数,也可以作为平面点 (a,b),也可以当作位置向量 $\overrightarrow{OA}$ $(A = (a,b))$,把它当

作形如

$$\begin{pmatrix} a & b \\ -b & a \end{pmatrix}$$

的二阶矩阵也无妨。即形如 $a+bi$ 的数，坐标为 (a,b) 的平面点，位置向量 $\overrightarrow{OA}(A=(a,b))$，二阶反对称矩阵

$$\begin{pmatrix} a & b \\ -b & a \end{pmatrix},$$

它们外形不同，但其实质均相同，即它们都是同构的，只要规定：

$$(a,b)+(c,d)=(a+c,b+d),$$

$$(a,b)\cdot(c,d)=(ac-bd,ad+bc),$$

$$\begin{pmatrix} a & b \\ -b & a \end{pmatrix}+\begin{pmatrix} c & d \\ -d & c \end{pmatrix}=\begin{pmatrix} a+c & b+d \\ -(b+d) & a+c \end{pmatrix},$$

$$\begin{pmatrix} a & b \\ -b & a \end{pmatrix}\begin{pmatrix} c & d \\ -d & c \end{pmatrix}=\begin{pmatrix} ac-bd & ad+bc \\ -(ad+bc) & ac-bd \end{pmatrix}。$$

同样，对向量，我们均不用向量积、数量积的方法定义，而采用两个位置向量按复数方式相乘。

复数的各种不同的同构形式，为求解问题带来了极大的方便，因为我们可以根据需要从一种形式转换为另一种形式，这常可大大简化问题的求解过程。

例如，证明 $\sum\limits_{k=1}^{5}\cos\dfrac{2k\pi}{5}$ 与 $\sum\limits_{k=1}^{5}\sin\dfrac{2k\pi}{5}$ 均为零。这即是证明

$$\sum_{k=1}^{5}\left(\cos\frac{2k\pi}{5}+i\sin\frac{2k\pi}{5}\right)=0。$$

该式中的复数若用与其同构的向量表示，则求解就很简便了。

设复数 $\cos\dfrac{2k\pi}{5}+i\sin\dfrac{2k\pi}{5}(k=1,2,\cdots,5)$ 在直角坐

标平面上对应的点为 $A_1, A_2, \cdots, A_5$，O 为坐标原点，且 $|OA_i| = 1 (i = 1,2,\cdots,5)$（图 3.21）。则问题化为证

$$\overrightarrow{OA_1} + \overrightarrow{OA_2} + \cdots + \overrightarrow{OA_5} = \overrightarrow{0}。$$

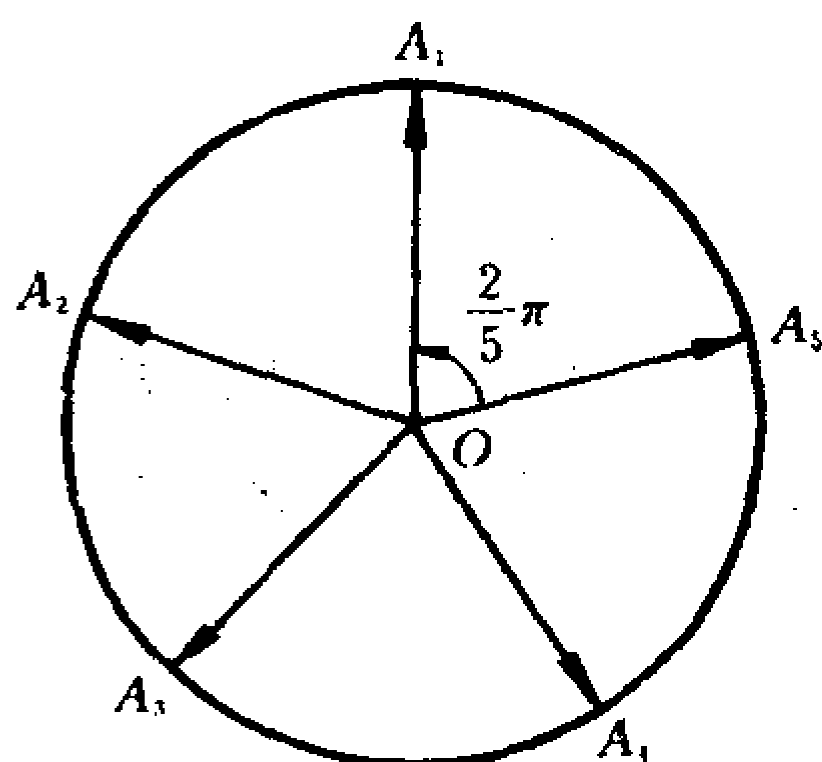


图 3.21

由于两相邻向量的夹角是 $\frac{2}{5}\pi$，故将每一向量逆时针旋转 $\frac{2}{5}\pi$ 后，五个向量的和必定不变。而要使一个向量旋转 $\frac{2}{5}\pi$ 后仍不变的必须是零向量，从而问题得证。

数学史上，更有应用同构方法使比较复杂又难以求解的问题变得简单的许多例子。其中，三大尺规作图不能问题，便是很有说服力的事例之一。

早在公元前 400 年左右，古希腊的巧辩学派就提出了三等分角、立方倍积和化圆为方三大尺规作图问题。为解决这些问题，历代数学工作者为之付出了巨大的努力，以至 19 世纪前，世界每个重要科研机关，都要收到数以千计的三大问题“解答者”的来信。为摆脱这种无休止的麻烦，1775 年巴黎科学院曾做过一项决议，不再审查三大难题的论文。直到近代，数学家们用同构的方法把它们转化为相应的代数问题后，才证明了这些问题是不可能仅用尺规作图来解决的。

比如，化圆为方问题，是将圆化为等积的正方形。其相应的代数问题是：用尺规作长为 $\sqrt{\pi}$ 的线段。由于尺规作图按形数对应的观点，无非就是利用直线与直线、直

线与圆、圆与圆相交等截取交点的几种基本方式来实施的，因而这些交点的坐标，相应地将是由两个一次、一个一次与一个二次、两个二次联立方程的代数解法来确定，它所能得出的数量都是由有理数经过有限次加、减、乘、除、开平方等运算表示出来的，而 $\sqrt{\pi}$ 不是这样的数量，故它不能用尺规作图来解决。可见，同构方法对处理问题的作用和威力。

例：一个实际问题的设计。

楼梯上有一盏电灯，问应该如何设计电路以使楼上与楼下均能自由开关它？

设楼下、楼上的开关分别记为 A、B，设计的电路就是一个新开关，记以 P。据设计要求，开关 A 或 B 翻转一次必须引起 P 翻转一次。

1° 设 $A = B = 1$ 时，$P = 1$。

据此，作出 A、B、P 的对应值表：

A	B	P
1	1	1
1	0	0
0	1	0
0	0	1

其方法是：第二行中，$A = 1$，$B = 0$，与第一行比较，知 B 从通到断，A 固定不变，而 P 亦应从 1 转到 0；第三行中，$A = 0$，$B = 1$，与第一行比较，A 从 1 变为 0，故 P 亦应从 1 变为 0，等等。由表知，A、B 在同一状态时 $P = 1$，在不同状态时，$P = 0$，借用命题代数语言有

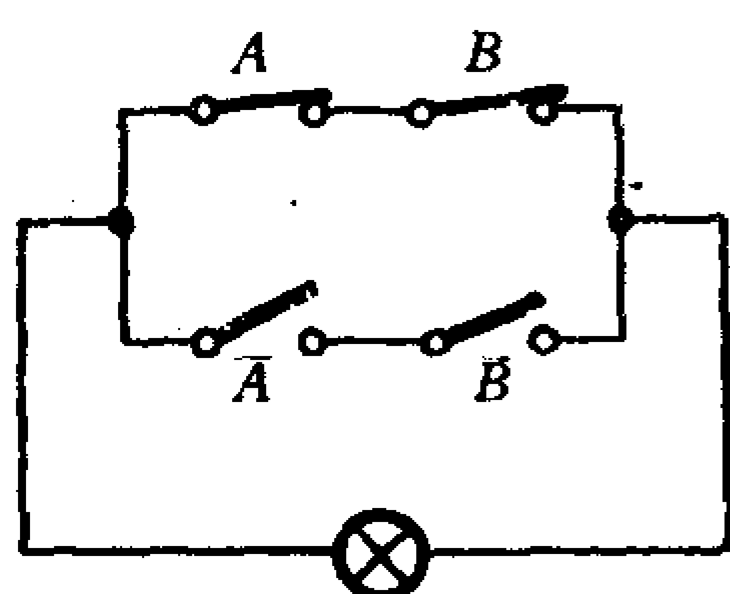
$$P = (A \land B) \lor (\overline{A} \land \overline{B})。$$

2° 设 $A = B = 1$ 时，$P = 0$。

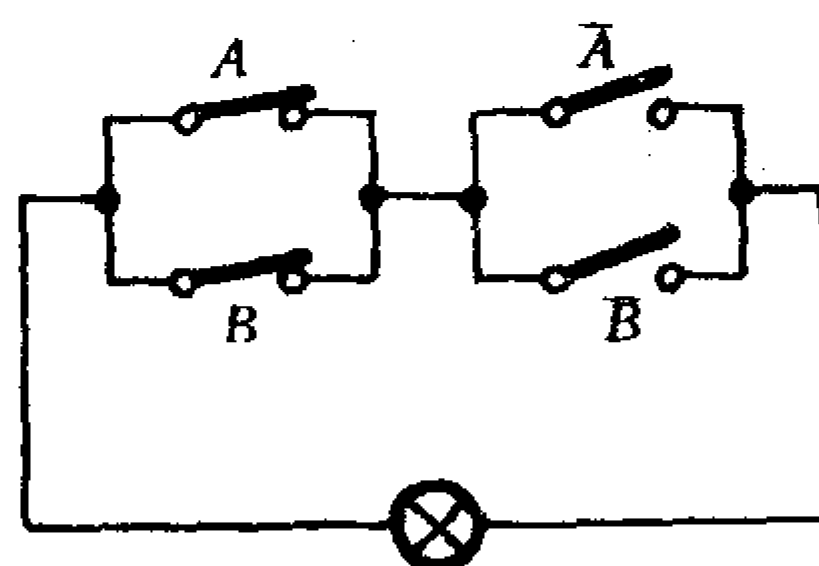
这时，P 与 1° 中的 P 处于反相，故

$$P=\overline{(A\wedge B)\vee(\bar{A}\wedge\bar{B})}$$
$$=\overline{(A\wedge B)}\wedge\overline{(\bar{A}\wedge\bar{B})}\quad（对偶关系）$$
$$=(\bar{A}\vee\bar{B})\wedge(A\vee B)。\quad（对偶关系）$$

其对应的电路图分别为图 3.22(1)（情况 1°）和 3.22(2)（情况 2°）。



(1)



(2)

图 3.22

参考文献

[1]田运,思维科学,浙江教育出版社,1988,p3,p14.

[2][苏]彼德罗夫斯基主编,普通心理学,人民教育出版社,p343.

[3]章志光主编,心理学,人民教育出版社,p.14.

[4][美]莫里兹编著,朱剑英编译,数学家言行录,江苏教育出版社,1990,p21,p68,p74,p93,p75,p84.

[5]邓东皋、孙小礼、张祖贵,数学与文化,北京大学出版社,1990,p43,p144,p267,p268,p40,p22,p140,p196—p197.

[6]恩格斯,自然辩证法,人民出版社,1971,p3,p235,p236.

[7]捷普洛夫,心理学,人民教育出版社,p197.

[8]张永声、钱锋、席振伟等,思维方法大全,江苏科学技术出版社,1990,p3.

[9][日]米山国藏,数学的精神思想和方法,四川教育出版社,1986,序.

[10]欧阳绛,数学方法溯源,江苏教育出版社,1991,p160,p133,p134,p16.

[11]高达声,略论科学符号,自然辩证法研究,1990,6(3),p27,p28.

[12]中国社会科学院主编,中国哲学史资料选辑(先秦之部),中华书局,1984,p566,p564.

[13]郑毓信、林曾,数学·逻辑与哲学,湖北人民出版社,1987,p122,p61.

[14][美]R.柯朗、H.罗宾斯著,汪浩、朱煜民译,数学是什么,湖南教育出版社,1985,p326.

[15][法]雅克·阿达玛著,陈植荫、肖奚安译,朱梧槚校,数学领

域中的发明心理学,江苏教育出版社,1990,p60.
[16][法]布尔巴基,数学研究者的数学基础,科学与哲学研究资料,1984(3);p24.
[17][美]M.克莱因著,张里京、张锦炎译,古今数学思想,上海科学技术出版社,1979年版第一册,p306,1981年版第四册p308,第一册p170,第四册p97—p98,1980年版第三册p344.
[18]张奠宙等著,近代数学教育史话,人民教育出版社,1990,p21.
[19]徐本顺、阴东升,对偶原理,曲阜师范大学学报(自然科学版),1990,16(1)、p106.
[20]徐利治,数学方法论选讲,华中工学院出版社,1983,p15.
[21]袁小明,世界著名数学家评传,江苏教育出版社,1990,p10—11,p154.
[22]B.R.盖尔鲍姆、J.M.H奥姆斯特德,分析中的反例,上海科学技术出版社,1980,p2.
[23]萧文强,数学证明,江苏教育出版社,1990,p107.
[24]朱梧槚,几何基础与数学基础,辽宁教育出版社,1987,p207.
[25][苏]A.Д.亚历山大洛夫等,数学——它的内容方法和意义,科学出版社,1984,p3.
[26][美]T.丹齐克,数,科学的语言(中译本),商务印书馆,1985,p16.
[27][英]李约瑟,中国科学技术史(数学),科学出版社,1975,p333,p32.
[28]李迪,中国数学史简编,辽宁人民出版社,1984,p3—p11.
[29]张奠宙等,现代数学与中学数学,上海教育出版社,1990.
[30]赵振威主编,中学数学教材教法(第一分册,总论),华东师大出版社,1990.
[31]朱梧槚、肖奚安,集合论导引,南京大学出版社,1991.
[32]席振伟、徐本顺,略论数学符号,曲阜师范大学学报(自然科学版),1992年第1期.
[33]席振伟,高等几何对中学几何教学的指导意义,曲阜师范大

学学报(自然科学版),1987年第2期.

[34]席振伟,略论对偶性与对偶思维方法,南京师范大学学报(自然科学版),1991年.

[35]席振伟、张明,论有向化的数学美,曲阜师范大学学报(自然科学版),1988年第4期.

[36]席振伟,怎样用射影几何观点研讨初等几何问题,赣南师范学院学报(自然科学版),1986年第2期.

[37]席振伟,刘徽数学思想研究,数学史研究文集,内蒙古大学出版社、九章出版社(台湾),1993年(第四辑).

[38]席振伟,中国古算符号论略,安徽师范大学学报(自然科学版),1993年.

［Ｇｅｎｅｒａｌ Ｉｎｆｏｒｍａｔｉｏｎ］
书名＝数学的思维方式
作者＝席振伟著
页数＝２１１
ＳＳ号＝１１４６０９８０
出版日期＝１９９５年１１月第１版